THE UNIVERSITY *of* EDINBURGH

Healthcare Costs and Employment Outcomes Associated with Cancer in Scotland

Kenneth Haining

Doctor of Philosophy

Deanery of Molecular, Genetic and Population Health Sciences

College of Medicine and Veterinary Medicine

University of Edinburgh

2023

Abstract

Background: Cancer has substantial economic costs which have received little study in the Scottish population. Scotland holds rich public healthcare data that can be linked to enable detailed measurement of the healthcare use of people with cancer over many years. Analysis of these data can enhance understanding of cancer costs and how they change over time.

Methods: I linked Scottish Morbidity Record (SMR) datasets and the Prescribing Information System (PIS) dataset to measure the healthcare use of people diagnosed with cancer and similar matched controls over eight years. Yearly and phase-of-care cost trajectories were charted. Generalised linear model (GLM) regression was used to analyse risk factors and estimate the excess costs of common cancers. I gained additional information by measuring costs for patients with and without pre-existing long-term conditions (LTCs). I also measured associations with employment outcomes in the United Kingdom Household Longitudinal Study (UKHLS) data using logistic regression and difference-in-differences (DiD) methods.

Results: Costs varied considerably by cancer site, with the highest total and excess eight-year mean costs measured in people with non-Hodgkin lymphoma. Trajectories of costs showed rates of cost accumulation were highest in the treatment and end-of-life phases, but more healthcare was used in total during the intervening continuing phase. I also found that patients with pre-existing LTCs used considerable healthcare, but the magnitude of association was greatest in the subgroup with no pre-existing morbidity, both as a ratio of baseline costs and in absolute monetary units. Further analyses found that cancer was significantly associated with reduced odds of working at 3–5 years after diagnosis.

Conclusion: This thesis brings new insights into the long-term economic costs of cancer, which will help policymakers, health economists and clinicians allocate healthcare resources more efficiently and better understand the economic burden of cancer.

Lay Summary

People with cancer use considerable amounts of healthcare such as hospital admissions, medicines and access to healthcare professionals. There may also be wider economic effects, such as loss of employment in working-age cancer survivors. However, little is known about the long-term economic costs of cancer in Scotland. In this thesis I measured the long-term healthcare use of people with cancer in Scotland and estimated the financial cost. I compared the healthcare use to that of people without cancer, to discover by how much cancer increased healthcare costs. I also measured the likelihood of working-age cancer survivors being employed compared to people without cancer.

I found that cancer patients used large amounts of healthcare in the year after diagnosis, and that healthcare use stayed elevated in subsequent years for survivors. The amounts of healthcare use varied considerably by the site of the cancer and how advanced it was at the time of diagnosis. However, more advanced cancers did not necessarily lead to more healthcare use. The highest healthcare use and costs tended to be found in cancers with moderate survival and, counter-intuitively, healthier cancer patients tended to use more healthcare overall. Furthermore, I found that cancer survivors were less likely to be in work than people without cancer up to five years after diagnosis.

An ageing society combined with improved cancer care will increase the number of people living with cancer, putting additional strain on healthcare and other services. My findings will help policymakers and health professionals allocate services efficiently, and enhance understanding of the long-term needs of people with cancer.

Declaration

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification except as specified.

Kenneth Haining

Acknowledgements

I would like to thank my supervisors, Dr Nazir Lone, Dr Peter Hall and Dr Elizabeth Lemmon, whose knowledge and guidance were so vital in helping me through this project.

I would also like to thank David Clark at eDRIS for his assistance in accessing the data upon which this project depended.

My thanks also, to my family, friends, colleagues and everyone else who supported me during this project.

List of Acronyms

A&E: Accident and emergency
ADRC: Administrative Data Research Centre
AHP: Allied health practitioner
AI: Artificial intelligence
AIC: Akaike information criterion
AME: Average marginal effect
ARDC: Australian Research Data Commons
ATT: Average treatment effect on the treated
AUC: Area under curve
BIC: Bayesian information criterion
BMI: Body mass index
BoE: Bank of England
BOCF: Baseline observation carried forward
CAPI: Computer-aided personal interview
CEVD: Cerebral vascular disease
CHEERS: Consolidated Health Economic Evaluation Reporting Standards
CHF: Congestive heart failure
CHI: Community health index
CI: Confidence interval
COI: Cost of illness
COPD: Chronic obstructive pulmonary disease
CPD: Chronic pulmonary disease
CRCD: Cancer-related cognitive dysfunction
CRD: Chronic respiratory disease
CRF: Cancer-related fatigue
CT: Computerised tomography
CVD: Cardiovascular disease
DALY: Disability-adjusted life year
DiD: Difference in differences
DRG: Diagnosis related group
DU: Discounted utility
eDRIS: electronic Data Research and Innovation Service
EOSC: European Open Science Cloud
ESRC: Economic and Social Research Council
FCM: Friction cost method
GFC: Global Financial Crisis

GLM: Generalised linear model
GP: General practitioner
HCM: Human capital method
HIV: Human immunodeficiency virus
HMRC: HM Revenue and Customs
HR: Hazard ratio
HRG: Healthcare resource group
HRQoL: Health related quality of life
ICD: International Classification of Diseases
ICDO: International Classification of Diseases for Oncology
ISD: Information Services Division
IV: Instrumental variable
KM: Kaplan-Meier
KMSA: Kaplan-Meier sample average
LLK: Lung, liver, kidney
LTC: Long-term condition
MCED: Multi-cancer early detection
ML: Maximum likelihood
MRI: Magnetic resonance imaging
MUP: Minimum unit pricing
NA: Not applicable
NHS: National Health Service
NICE: The National Institute for Health and Care Excellence
NMSC: Non-melanoma skin cancers
NRS: National Records of Scotland
NZ: New Zealand
OECD: Organisation for Economic Co-operation and Development
OLS: Ordinary least squares
OR: Odds ratio
P&CFS: Practitioner and Counter Fraud Services Division
PAF: Population-attributable fraction
PET: Positron emission tomography
PIS: Prescribing Information System
PLICS: Patient Level Information and Costing Systems
POC: Phase of care
PPI: Critical Care Patient-Public Involvement in Research
PPP: Purchasing power parity

PUD: Peptic ulcer disease
PVD: Peripheral vascular disease
QALY: Quality-adjusted life year
QoL: Quality of life
Q-Q: Quantile-quantile
RECORD: REporting of studies Conducted using Observational Routinely-collected Data
ROC: Receiver operating characteristic
RR: Relative risk
SD: Standard deviation
SF: Short form
SIMD: Scottish Index of Multiple Deprivation
SMR: Scottish morbidity record
SOCRATES: Scottish Open Cancer Registration And Tumour Enumeration System
SPA: State Pension Age
TNM: Tumour, node, metastasis
TRE: Trusted research environment
UK: United Kingdom
UKHLS: United Kingdom Household Longitudinal Study
US: United States
USD: United States Dollars
WTP: Willingness to pay
YPLL: Years of productive life lost

Contents

List of Figures

List of Tables

1 Introduction

1.1 Background

Cancer has been reported as the leading cause of death in 28 out of 40 European countries and the second largest in the remaining 12 [1], with approximately two million deaths and four million new cases annually [2]. Worldwide, more than eight million people die from the disease each year [3]. Yet recent decades have seen major advances in the detection and treatment of cancer, with survival rates in the United Kingdom (UK) doubling over the last 40 years [4]. Although there is little optimism that a cure will be found in the foreseeable future, it has been suggested that cancer will be turned into a long-term chronic disease such as diabetes or asthma, which patients will manage through personalised medications [5, 6].

Despite, or even because of, improving outcomes, the costs of cancer to patients and wider society are rising [7, 6]. Global spending on cancer medicines reached almost US$150 billion in 2018, representing approximately 10% of the $1.5 trillion global spending on all medicines, after rising more than 10% each year from 2013. In the United States (US), cancer spending more than doubled during this period to reach $57 billion, while annual world spending on oncology is projected to be in the range $220–250 billion by 2023[1] [8]. As this estimate only includes direct oncology spending, total cancer-related spending is likely to be higher still. This follows a general pattern of rising health spending in high-income countries; in the European Union (EU) cancer's share of total health expenditure was approximately stable during the period 1995–2014, despite an inflation-adjusted increase in total spending of almost 50% (€50.5 to €83.2 billion in 2014 prices), while production losses due to cancer-related mortality were found to have decreased 11% from €54.5 billion to €48.6 billion, primarily due to a reduction in lung cancer for 40–60 year old men [9]. During this period, cancer spending in the UK was estimated to have increased from €4.5 billion to €9.5 billion, however, the 2014 figure represented a decrease from €10.3 billion in 2005 [9].

Before setting out the aims and context of this thesis I shall describe current trends in the factors driving changes in cancer incidence, survival and expenditure, as understanding these will help to frame the epidemiological and socioeconomic contexts for this thesis.

[1] all in United States Dollars (USD)

1.1.1 Risk Factors

Increasing affluence in high-income countries has been accompanied by rising rates of cardiovascular disease, cancer, respiratory diseases and diabetes, while the prevalence of these diseases is now growing even more rapidly in low and middle-income countries [10]. Examining the trends in underlying risk factors will aid understanding of trends in cancer incidence and survival, and how these impact healthcare use and associated costs. The risk factors described here were identified by Public Health Scotland [11] as important ones in Scotland, and that were associated with a range of cancers, rather than a single cancer or multiple cancers affecting a particular body area. I also describe socioeconomic status as evidence suggests an association with cancer that is not fully explained by differences in smoking, alcohol intake and diet [12]. Other risk factors, such as the age at having a first child for female breast cancer [11], were not described as their association was with a particular cancer or body area.

Age Demographics

Life expectancy at birth increased steadily from 68.5 years in 1960 to 80.7 years in 2018 across high-income countries [13]. Increases in UK life expectancy, in combination with a decrease in fertility, are driving an ageing population which is increasing the demand for health services and social care [14]. In England, the proportion of people over 65 years old is projected to increase from 18.2% in 2018 to 20.7% in 2028 [15], and to around 25% by 2050 [16], by which time the percentage over 80 is expected to more than double to 10.3% [16]. In Scotland, the percentage over 65 years old is projected to increase by 4% during the period from mid-2018 to mid-2028, while the percentage of those over 80 is expected to increase by 25% [17]. Although longevity has increased for UK citizens, the years of life gained have not necessarily been healthy ones, with healthy life expectancy of 63.1 years for UK men and 63.6 for UK women during 2016–19. This was lower in Scotland with 61.9 for men and 61.8 for women. Men in Scotland have the lowest life expectancy in the UK at 77.1 years, but spend a greater proportion (80.3%) of their lives in good health. However, the proportion of life spent in good health decreased from 79.9% to 79.5% for men in the UK during 2016–19 and from 77.4% to 76.7% for women [18].

Tobacco Smoking

A major risk factor for several cancers, the prevalence of tobacco smoking has been declining, with the proportion of UK adults who are smokers falling from 20.2% in 2011 to 14.1% in 2019 [19]. Smoking is more prevalent in men, 15.9% being smokers

compared to 12.5% of women. Scotland has a higher proportion of smokers than the rest of the UK, with 23.4% of adults being smokers in 2011 and 15.4% in 2019 [19]. These data should be interpreted cautiously due to being self-reported, but show a pattern of declining smoking similar to other high-income countries. Scotland's government aims to reduce prevalence to below 5% by 2034, however, it is unlikely that this target will be achieved [20]. Due to its association with cancer, the effect of smoking on healthcare use and associated costs is considerable; for example, lung cancer alone has been found to account for about a fifth of the total costs of cancer for Medicare in the US [21]. Around one-quarter of cancer deaths in Scotland are attributed to lung cancer, more than double the number of deaths from colorectal cancer, the second largest cause of cancer deaths [22]. Smoking has been noted as a driver of health inequalities in cancer incidence and mortality [23]. It is more common in deprived areas with prevalence rates of approximately 35%, compared to 10% in the least deprived areas, and the rates of lung cancer in deprived areas are approximately three times higher than those in affluent areas [24]. The strong association between social class and cancer outcomes has prompted claims that removing social class disparities would reduce cancer mortality more than innovative treatments [25].

Adiposity

While tobacco smoking has declined, adiposity rates have been increasing, with attributable deaths in England and Scotland rising from 18% in 2003 to 23% in 2017, making adiposity a larger contributor to mortality than smoking [26]. The National Health Service (NHS) in England recorded more than one million admissions related to obesity and being overweight during 2019/20, an increase of 17% over 2018/19 although the increase may be due to changes in the recording method, and the overall trend over the previous ten years was approximately flat [27]. A 2019 survey found that 28.0% of adults in England were obese and 36.2% overweight but not obese, with men around 8% more likely to have excess weight than women, and residents of the most deprived areas around 9% more likely to have excess weight than those in the least deprived areas [28].

Diet

Cancer costs caused by poor diet were reported to have risen over 50% from 4.0% of the total NHS budget for the UK during 1992/3, to 6.2% during 2006/7. But overall costs arising from poor diet fell overall, from 26.0% to 21.6% over the same period [29]. A more recent study found little change in the UK diet over the period from

2008 to 2017, with all groups based on age and sex consuming less than the recommended daily intake of fruit and vegetables. However, the intake of free sugars, and of red meat and processed meat showed slight downward trends [30]. In Scotland, the period between 2001 and 2015 saw little change in the diets of Scottish people and most dietary targets were not met. The mean food energy density rose, although there was no change in total fat intake and free sugar and saturated fat intake fell slightly. There was little change in fruit and vegetable consumption with deprived households consuming less than less deprived households [31]. In 2021, only 22% of adults in Scotland consumed the recommended amount of fruit and vegetables, with little change in the figure since 2003. Also in 2021, only 20% of adults in Scotland met the recommendations for energy density and only 6% ate the recommended intake of fibre [32].

Physical Inactivity

Cancer costs from physical inactivity as a proportion of total NHS costs in the UK were reported to be higher during 2006/7 than during 1992/3, primarily as a result of a doubling of the proportion taken by breast cancer from 0.3% to 0.6%. However, overall costs from physical inactivity fell from 8.5% to 6.5% during the same period on account of a decrease in the proportion of costs attributed to ischaemic stroke [29]. In a more recent study, two thirds of Scottish adults were reported to meet the guidelines for physical activity, however this dropped to 54% for people in the most deprived areas and women were less likely than men to have met the guidelines [11]. Levels of physical activity have been rising in Scotland: the percentage of adults meeting the recommended levels of moderate or vigorous physical activity (MVPA) rose from 62% in 2012 to 69% in 2021, with a higher proportion of men (73%) than women (65%) meeting the recommended levels in 2021 [32].

Alcohol Intake

While it has been reported that moderate levels can have beneficial effects on health [33], alcohol intake is believed to be a risk factor for cancer at all levels of intake [11]. Of the 6.5% of Scotland's deaths attributed to alcohol, 28% are thought to be caused by cancer [34]. The costs of cancers attributed to alcohol use were reported to have risen from 4.0% of total NHS costs in the UK during 1992/3 to 6.0% during 2006/7 [29] However, alcohol consumption is also reported to have declined over the past fifteen years in Scotland, England and Wales from highs in 2004. Although the declines were greatest in Scotland, consumption in 2020 remained higher than in

England and Wales (9.4 litres of pure alcohol sold per adult in Scotland compared to 8.8 litres per adult in England and Wales) [35]. The proportion of people in Scotland, England and Wales who do not drink alcohol rose from 18.8% in 2005 to 20.4% in 2017. The proportion of people who drank moderately was lower in Scotland (53.5%) than in England (57.8%) but higher than in Wales (50.0%). However, the proportion of people who drank more than 6–8 units was higher in Scotland at (37.3%) than England (26.2%) and Wales (30.4%) [36]. Young people aged 16 to 24 were less likely to drink than people aged 45 to 64, but people aged 65 and over were most likely to drink. High earning professionals and managers were most likely to drink (69.5%) while routine and manual workers were least likely (51.2%) [36]. In Scotland the mean units of alcohol consumed per week among all adults declined from 16.1 units in 2003 to 11.3 units in 2021. There was considerable differences between the sexes; men drank almost twice as much as women (14.8 units compared to 8.0 units in 2021). The rate of hazardous/harmful drinking declined from 34% in 2003 to 25% in 2013 and has stayed around this level to the time of the report in 2021 [32]. Hazardous drinking in men in Scotland declined almost continuously from nearly 50% in 2003 to around 31% in 2021. The gap between men and women narrowed over this time period, however even in 2021 the proportion of men was almost twice as high as women at 16% [32].

Socioeconomic Status

The strong association between social class and cancer outcomes has led to claims that removing social class disparities would reduce cancer mortality more than innovative treatments [25]. This association may be rooted in the association between smoking and lower socioeconomic status. Rates of smoking are declining in the UK [19] but against these declines must be set the increases in other cancer risks such as overweight and obesity, which are also associated with lower socioeconomic status [12]. Additionally, an association between cancer and lower socioeconomic status has been found that was not explained by differences in smoking, diet and exercise [12]. Socioeconomic differences may be amplified by screening programmes as these can have lower take-up in people with lower socioeconomic status [37].

1.1.2 Cancer Incidence

Trends in UK cancer incidence have been found to differ from those of similarly developed countries, with increasing incidence and decreasing mortality seen for certain cancers compared to decreasing incidence and decreasing mortality for the same cancers in other high-income countries [38]; for example, in the US, incidence

has declined from a peak of around 500 per 100,000 in the 1990s to around 450 per 100,000 in 2017, although it rose slightly for women over this period [39] and heterogeneity among cancer sites was reported; cancers of the lung, prostate, stomach, larynx, and oesophagus saw declines in incidence whereas cancers of the breast, kidney, skin, liver, testis and pharynx saw rises [40]. In Scotland, the incidence rates of all cancers rose during the period 1994–2018 but decreased by 3.5% for men and women combined during 2008–2018, although this figure was estimated [24]. However, the total number of cancers increased from 30,600 in 2009 to 34,000 in 2018, due to the rising overall number of elderly individuals. Again, variations between cancer types were recorded, with decreases during 2008-18 seen in lung cancer (-10.3%), colorectal cancer (-18.1%) and oesophagus cancer (-6.9%), while increases were seen in skin cancer (+7.0%), kidney cancer (+17.8%), breast cancer (+1.5%) and head and neck cancer (+2.4%). [24]. The combined figures for all persons disguise discrepancies between sexes; lung cancer rose 1.5% for women and fell 18.7% for men, while kidney cancer rose 25.9% for men and skin cancer 14.8% [24]. The rise in lung cancer rates in women is believed to be caused by women taking up smoking later than men during previous decades and is reported as likely to reverse [24]. Despite the variation in incidence across countries, some broadly similar patterns can be seen, for example, breast, prostate, lung and colorectal cancers have the highest incidence in the UK, accounting for more than half of all registrations, and are also the most common cancers worldwide [37].

1.1.3 Screening

The effects of screening on overall incidence and prevalence are unclear and may vary by cancer site: a systematic review found screening raised the incidence of prostate cancer [41], while a systematic review of colorectal cancer found incidence was reduced [42]. Other factors that may affect the impact of screening on incidence are the population being screened, the healthcare system and implementation of the screening programme itself [23]. It has been suggested that increases in incidence may occur soon after the screening programme is implemented but gradually reduce as people with cancer die at faster rates than people without cancer [23]. Screening may increase inequalities in cancer outcomes due to lower take up by people with lower socioeconomic status [37, 23].

1.1.4 Technology

New technologies drive much of the increase in cancer spending, as they tend to be more expensive and also make treatment available to more patients, which in turn drives overall costs upward [7]. Increased cancer screening can also increase expenditures, due both to screening costs and to an increased number of survivors who require ongoing care and surveillance [7], while the treatment of once incurable cancers causes a constant increase in the economic burden of care [43]. Despite improving outcomes, a lack of an association between treatment costs and clinical benefits has been reported [44].

Anger et al. (2019) identified eight key factors driving oncology trends [8]:

1. digital health technologies
2. increased focus on patient-reported outcomes (PROs)
3. real-world data
4. predictive analytics and artificial intelligence (AI)
5. shifts in drug types
6. increased availability and ease of biomarker testing
7. availability of pre-screened patient pools and recruitment
8. changes in the regulatory landscape.

This list highlights the importance of new technologies in driving cancer trends, but also shows that medicines are only part of the picture. With regard to cancer technology spending, however, drug development is a key factor. New drug brands had a median cost of US$148,800 per patient treatment year in 2018, though this figure was a slight decrease from US$150,000 in 2014 [8]. The primary goals driving pharmaceutical companies to develop cancer drugs are said to be improvements in the quality of patients' lives, and the desire to earn profits [45]. Although the first of these goals has brought life-extending treatments, this has often come at a high cost for individuals and healthcare systems, with some treatments unaffordable for those who need them most [45]. While new drugs tend to be more expensive, their effectiveness may lead to reductions in costs elsewhere, resulting in a greater proportion of total expenditure accruing to drugs [7]. The cost of new drugs can strain public healthcare systems, which can lead governments to cut spending in other areas such as infrastructure and education, and may lead to increased rationing in healthcare [45].

This effect is likely to increase if cancer is made a controllable, chronic disease through life-extending drugs, because the prevalence of cancer will increase with a consequent increase in costs and the need for social care facilities [6].

A disproportionately high level of UK research expenditure has been channelled into cancer compared to other diseases, relative to disease burden [46], though the difference may be declining [47]. Oncology has been reported to represent 47% of Phase I clinical trials in 2018 out of nine key therapy areas [8]. Research spending on cancer is highly geared towards new drugs and other treatment technologies, with only a small proportion spent on prevention [48]. Redressing this imbalance has been stressed as a major challenge for cancer care [23]. Factors that could reduce future costs include better predictive and prognostic factors to enable more precise targeting of therapies [7]. Genetic tests and biomarkers could identify those patients who will benefit from new drugs, compensating for the higher costs of new drugs [7]; clinical trials involving biomarker-based stratification represented 39% of oncology trials during 2018 [8]. Furthermore, oncology drugs can take considerable time to reach patients, with a median time of 10.5 years reported from the time of the first patent filing, and most medicines only reaching patients in a small number of high-income countries [8]. On top of these factors is the impact of Covid-19, which may affect the progress of cancer research for many years [49].

1.1.5 Survival and Prevalence

The risk of dying from cancer has been reported to have fallen by around 10% in the UK between 2009 and 2018, despite increasing incidence and despite cancer being the most common cause of death in England [37]. UK cancer survival rates have doubled over the last 40 years [39], following a pattern of decreasing mortality and increasing survival for cancer and other non-communicable diseases in most high-income countries: between 2000 and 2019 mortality from non-communicable diseases cardiovascular disease (CVD), cancer, diabetes and chronic respiratory disease (CRD) fell from 16.5% to 10.3% for people aged 30–70 years in the UK and from 16.6% to 11.9% for all high-income countries [50]. It should be noted, however, that although mortality rates have decreased, the overall number of UK deaths from cancer is increasing annually due to a rise in the total number of older people [22]. In Scotland, mortality rates for cancer patients declined 13% for men and 7% for women in the ten years prior to 2016. Despite these improvements, five-year relative survival in England, Wales and Scotland is lower than the European average for men and women [4]. Mortality rates for all causes in Scotland are the highest in western Europe, partly as a

result of poor diet in some regions that leads to high frequencies of cardiovascular disease and cancer in older adults [51]. Statistics on cancer mortality should be treated with caution, as death data are generally less accurate than registry statistics, due to difficulties in correctly attributing the cause of death [52].

Longer survival after a cancer diagnosis has, *ceteris paribus*, the effect of increasing prevalence of cancer in the population. Improvements in detection and treatment, together with a higher overall incidence due to ageing populations, can be expected to increase the prevalence of some cancers. Cancers with high prevalence tend to be those that combine high incidence with high survival [53]. In Scotland during 2015, the cancers with the highest prevalence for people over 65 years old were breast cancer for women and prostate cancer for men [53]. Although the general trend is of increasing cancer prevalence the full picture is more complex. In high-income countries, the increase in cancer incidence is driven primarily by ageing societies, while the increase in survival is driven mainly by improved treatment and detection. Incidence of lung cancer is falling, which is likely to increase the relative incidence of other cancers with better survival, such as breast cancer, prostate cancer and colorectal cancer, thus moving cancer to a more long-term disease when all sites are considered together. During 2010–11, 50% of cancer patients survived for at least ten years after diagnosis [4]. Lung cancer has been reported as having the highest total cost to society, due to high incidence and high mortality leading to considerable use of healthcare resources [54]. However, all individuals must ultimately die from something and the end-of-life period is known to be costly for other conditions [55], so part of lung cancer's cost may only be delayed rather than eliminated as incidence falls. Furthermore, the increase in longevity will incur societal costs as survivors continue to use healthcare and other services. If individuals who might have perished from lung cancer survive with long-term conditions like obesity, diabetes, dementia, and other diseases that share risk factors of lung cancer such as social deprivation, it can be expected that future demand for healthcare, social care and other services will be affected.

1.1.6 Scotland's Population

Scotland's population has poorer health [56] and lower life expectancy than other western European nations [51, 57, 16] due to a complex mix of health behaviours and socioeconomic factors that increase risk of common diseases [57, 56, 58]. The mortality rates of lung cancer and heart disease in Scotland are the highest in western Europe [59] while overall cancer mortality rates are higher than in other European nations [51], with a hazard ratio 1.4 times that of England [60]. Almost half (44%) of

Scottish adults have at least one long-term health condition [16]. The prevalence of smoking in Scotland is higher than in other high-income countries, leading to public health policies to reduce prevalence to below 5% by 2034. Alcohol consumption is also high; of Scotland's 57,327 deaths in 2015, 3,705 (6.5%) were attributed to alcohol. Cancer accounted for more than one quarter (28%) of these alcohol-attributed deaths. Scotland is also perceived to have a particularly unhealthy diet [59] due to factors such as sociodemographic deprivation [58] and cultural perceptions regarding healthy eating [61].

1.1.7 Summary

Sections 1.1.1 to 1.1.5 describe an overall picture of changing prevalence of risk factors with improving treatment and detection. The decline in the prevalence of tobacco smoking is reducing the relative incidence of lung-related cancers, but ageing populations and increasing adiposity are raising the incidence of other cancers with generally lower mortality. Together with improved detection and treatment, these trends are likely to improve cancer survival and make cancer more of a long-term chronic disease. This is likely to increase the economic costs costs of cancer to patients and societies, and has implications for how these costs should be measured. This could be particularly important in Scotland due to the relatively poor health of the population and the high, though declining, prevalence of tobacco smoking.

1.2 Aims and Objectives

To efficiently allocate healthcare and other services, healthcare providers and policymakers will need to take into account the trend towards cancer as a long-term condition. Currently, the long-term costs of cancer are not well understood, particularly in Scotland. In this thesis my aim was to enhance understanding of the long-term economic costs of cancer by reporting healthcare use and the associated costs of people with cancer in the Scottish population, using linked administrative data. The following specific research questions were posed.

1. How much healthcare do cancer patients in Scotland use and what are the associated monetary costs?
2. How do the costs vary over time?
3. Which factors influence costs and what is their relationship to survival?
4. How do costs compare to similar individuals without cancer?

5. What is the relationship between cancer and costs for patients with long-term conditions (LTCs)?
6. How does cancer affect other socioeconomic outcomes such as employment?

These questions gave rise to the following objectives.

1. Measure the healthcare use of cancer patients in Scotland and their associated costs.
2. Chart trajectories of costs over time.
3. Identify risk factors of costs and compare them to risk factors of survival.
4. Compare costs to patients without cancer and estimate excess costs.
5. Measure costs for patients with LTCS.
6. Measure the associations of cancer with wider socioeconomic outcomes such as employment.

1.3 General Approach

To meet these objectives and answer the research questions, I undertook retrospective analyses using linked public-sector administrative datasets. A considerable proportion of work for the thesis was spent on researching, applying for and linking public datasets. Where such datasets were not available to answer a particular research question, I took evidence from publicly available survey data. Reporting of the analyses followed recommendations from the REporting of studies Conducted using Observational Routinely-collected Data (RECORD) and relevant components of the Consolidated Health Economic Evaluation Reporting Standards (CHEERS).

As will be discussed in Section 2.3.10, considerable heterogeneity in methods, research questions and populations makes systematically reviewing the costs of cancer challenging. As the literature on cancer costs is extensive, a systematic review was considered infeasible. Literature results related to specific analyses are considered in the relevant chapters.

1.4 Thesis Structure

1.4.1 Chapter 2: Measuring the Changing Costs of Cancer: A Literature Review of Measurement Methods

Chapter 2 contains a review of the literature on methodologies around measuring the healthcare costs of cancer. I describe the various approaches taken in costing studies and how the different methodologies can impact results. I discuss how the considerable heterogeneity in methods, populations and health systems makes generalising results from other studies problematic and consider practical solutions. This provides the rationale for the methodologies guiding future chapters.

Aims and Objectives of Chapter 2

The aim of Chapter 2 was to understand the methods of measuring cancer costs and their relationship to the cancer trajectory. The objectives were as follows.

1. Describe the methods of measuring cancer costs.
2. Discuss strengths and limitations of the methods.
3. Frame the approaches of analyses that follow in the context of the wider literature.

1.4.2 Chapter 3: Eight-Year Healthcare Resource Use of Cancer Patients in Scotland Using Linked NHS Datasets

Chapter 3 describes how I met objectives 1–3 by linking several NHS Scotland datasets to measure the healthcare use of Scottish cancer patients over time. Eight-year incidence costs for each of the ten most common cancer types in Scotland were estimated using per-episode costing with unit costs derived from the Scottish Costs Book. Trajectories of costs were charted by year and by phases of care. Risk factors for costs were determined using generalised linear model (GLM) regression and compared to risk factors of mortality derived from Cox regression.

Aims and Objectives of Chapter 3

The overall aim was to increase understanding of the economic costs of cancer by measuring the long-term healthcare use of cancer patients in Scotland. An additional aim was to better understand the factors driving the costs of healthcare use and whether the same factors that influence survival also influence costs. The analysis had the following objectives.

1. Describe the dynamics of healthcare costs and survival by charting trajectories..
2. Examine how costs vary between different cancer types.
3. Measure and compare risk factors for costs and survival.
4. Explore the potential of linking public-sector datasets to gain a wider picture of costs.

Overarching Methods

The datasets used to measure resource use were common to the analyses described in chapters 3 and 4 and, to some extent, in Chapter 5. The methods for assigning costs were also the same across these analyses. Therefore, the methods in Chapter 3 that describe the datasets and cost assignment can be read as overarching methods for these chapters. The description of variables in this chapter is also very relevant to chapters 4 and 5, but with some differences that are dealt with in the relevant chapters.

1.4.3 Chapter 4: Comparison with a Matched Control Group

Chapter 4 details how I tackled Objective 4 and extended the results of Chapter 3, comparing the healthcare costs of cancer patients with a matched control group. I gained further insight into the impact of cancer on costs by measuring the healthcare use and associated costs of patients with and without cancer, charting the trajectories over time, and calculating excess costs over the eight-year period.

Aims and Objectives of Chapter 4

The overall aim was to better understand how the healthcare use and associated costs of a cancer cohort compared to a similar cohort without cancer, with emphasis on trends in costs over time. The analysis had the following objectives.

1. Measure the long-term excess costs of cancer in the Scottish population.
2. Chart and compare trajectories of resource usage and associated costs over time for people with and without cancer.
3. Determine when, if at all, mean excess costs reverted to zero after the cancer diagnosis.

1.4.4 Chapter 5: Inpatient Cost Trajectories For Patients With Long-Term Conditions

Chapter 5 describes the measurement of inpatient use and associated costs over time for patients with underlying long-term conditions (LTCs). This relates to Objective 5. I also measured the associations of cancer diagnoses with costs specific to the LTCs and charted cost trajectories.

Aims and Objectives of Chapter 5

The overall aim of Chapter 5 was to better understand the costs of cancer for patients with underlying LTCs. Specific objectives were as follows.

1. Measure and chart cost trajectories of cancer patients with pre-existing LTCs, and similar individuals with the same LTCs but no cancer diagnosis.
2. Measure the association of a cancer diagnosis with healthcare costs for patients with LTCs.
3. Measure the association of a cancer diagnosis with costs specific to the LTCs.

1.4.5 Chapter 6: Long-Term Employment Outcomes for Cancer Survivors in UKHLS

Chapter 6 addresses Objective 6 by investigating the association of a cancer diagnosis with employment outcomes, using the UK Longitudinal Survey (UKHLS) dataset. Cross-sectional data were analysed using logistic and linear regression for cancers occurring prior to the survey period. To investigate a causal relationship, longitudinal data were analysed using difference in differences (DiD) methods.

Aims and Objectives of Chapter 6

The overall aim was to better understand the relationship between a cancer diagnosis and work-related outcomes for cancer survivors in the UK, with emphasis on whether changes in outcomes persist beyond the treatment period. Specific objectives were as follows.

1. Measure the long-term association between a cancer diagnosis and being in paid work.
2. Measure the long-term associations between a cancer diagnosis and other work-related outcomes such as earnings and income.

1.4.6 Chapter 7: Overall Discussion

Chapter 7 summarises the results from previous chapters and attempts to synthesise them to provide an overall picture that meets the aims of the thesis. Findings of the results taken together are described and the overall strengths and limitations are detailed. The overall findings are then discussed in light of the overall limitations and their external validity to other populations considered. I conclude by discussing the implications for policy and future research.

Aims and Objectives of Chapter 7
The aim was to consider how the findings from the analyses answered the overall aims of the thesis. The objectives were as follows.

1. Summarise the previous analyses.
2. Consider how the results, taken together, answered the overall aims and research questions of the thesis.
3. Discuss overall limitations, interpretation and policy implications.

1.5 Challenges

A substantial part of this thesis involved the data acquisition process. My initial interest in the project had been based on studying socioeconomic outcomes such as changes in employment status, with healthcare use of lesser interest. Additionally, the timescale in my project proposal was based on having access to data early in the project. It became apparent that the data acquisition process was likely to take considerable time, and that data acquisition on non-healthcare outcomes would have additional burdens. As my project funding was only for three years it was necessary to apply for data at the beginning of the project to have a reasonable chance of accessing data within the PhD time frame. During the data acquisition process the Administrative Data Research Centre (ADRC), which had been processing my application, shut down and I was assigned a new data contact within the electronic Data Research and Innovation Service (eDRIS), which acted as ADRC's successor. This delayed the application process considerably as I essentially had to begin the application process again with a new data contact. The uncertainty over which data would be available to analyse caused difficulties in deciding what the focus and specific research questions of the thesis should be. This uncertainty, combined with the large body of literature related to cancer costs, made reviewing the literature challenging, and it was necessary to take a broad approach. As my interest was in employment

outcomes I located publicly available data to analyse relevant data, the results of which are described in Chapter 6. This also led to the addition of health economist Elizabeth Lemmon to the supervisory team, whose expertise contributed greatly. It seemed unlikely that analysis of survey data alone would be substantial enough for a PhD. Therefore I switched to part-time study while taking employment with a health informatics company, and moved the focus of analysis to healthcare use when that data became available. Combining work and study was an additional challenge but one that perhaps brought benefits as it forced me to prioritise areas of study and organise my time more efficiently.

Compounding the problem of accessing data was the impact of the Covid epidemic. During this time my main supervisor was fully occupied with intensive care duties for many months, meaning I was unable to call upon his expertise and guidance. There was also considerable disruption to data access and other services, and working from home brought its own challenges. I mitigated these problems as best I could with support from my additional supervisors and was grateful to receive a funded extension of my studies. Further challenges came from the broad scope of the project, covering epidemiology, data science, health economics and social economics. Coming from a background in information technology and maths, I had to undertake a considerable amount of learning in other areas, and understand the literature of multiple disciplines. The expertise and endless patience of my supervisory team were critical in guiding me through these and other challenges.

2 Measuring the Changing Costs of Cancer: A Literature Review of Measurement Methods

2.1 Introduction

In Chapter 1, I described cancer trends that are likely to impact on the economic costs of cancer. The review in this chapter will provide the methodological context for the analyses to follow, in light of the trends described in Chapter 1. The objectives were to describe the methods of measuring cancer costs, discuss their strengths and limitations, and frame the thesis and the approaches of my analyses in the context of the wider literature. I will first examine what drives costs over the trajectory of a cancer patient, as this will provide useful background context for understanding the measurement of costs. Following this, I will describe how costs are measured in the literature and consider the continuing appropriateness of existing methods. The opportunities offered by increasing amounts of administrative data will be examined, and also associated issues such as those relating to privacy, security and processing. Problems with existing methodologies in costing and how they cause heterogeneous results will be explored. I shall end with a discussion of what the described trends imply for future cancer costs and how those costs can be measured. My focus will be on the Scottish population and that of the wider UK. Other populations will be considered for context and if they are likely to foreshadow UK trends, as is believed to be the case for the US [6]. The evidence in this chapter was identified using Medline and Embase, Web of Science, Google Scholar and expert knowledge. Example search terms were: (cancer AND (research OR R?&?D) AND (spending OR expenditure OR cost?)), (trend* AND (smoking OR tobacco) AND Scotland OR UK OR United Kingdom OR Britain). The websites of relevant national statistical agences, national health systems and cancer organisations were also used where these were expected to be unbiased. Systematic reviews were used where possible however most articles were not evidence-based.

2.2 Cost Drivers on the Patient Trajectory

2.2.1 Pre-diagnosis

While studies tend to focus on cancer costs after diagnosis, the economic costs of the disease may begin even before this with increased use of primary care and other healthcare such as cancer screening. Patients may require time off work to access healthcare and may also incur costs for transport, parking, and other services. Patients, relatives and employers may incur costs from morbidity-related productivity. Healthcare providers and wider society can incur costs for awareness-raising health

campaigns and other preventive measures, although spending on disease prevention is generally low compared to other healthcare spending [62].

2.2.2 Diagnosis and Treatment

Substantial healthcare resources may be required to provide a cancer diagnosis and to determine the nature and extent of the tumour. In addition to laboratory costs, there may be substantial up-front costs for sophisticated diagnostic technologies such as computerised tomography (CT), magnetic resonance imaging (MRI) and positron emission tomography (PET), as well as ongoing costs for skilled operators and other oncology professionals [6]. Once cancer is diagnosed, considerable healthcare use tends to occur during the six months after diagnosis [63], when patients may undergo curative treatment such as surgery and radiotherapy, or palliative care for those with poorer prognosis. Since the 1980s there has been a shift away from hospital costs towards drug costs, particularly in the US; however, inpatient costs remain a major portion of total cancer costs [6, 64]. Major determinants of costs include stage of diagnosis [63] and intensity of treatment [21]. High-quality palliative treatment can lead to better survival, but can incur unsustainable costs [65]. More than a third (38%) of patients undergo chemotherapy or other life-sustaining treatments in the last month of life [66].

2.2.3 Mortality

Resource use has been observed to increase exponentially as death approaches [67, 66], particularly in the final three months of life, with hospital admissions comprising the largest proportion of costs in this period [66]. During the last month of life many cancer patients use chemotherapy or other life-sustaining treatments [66]. Uncertainty surrounds the economic costs arising from mortality. Outlays accruing to relatives such as funeral costs tend not to be counted, while productivity losses due to premature mortality have been estimated and found to be substantial and increasing for cancer in many different countries [3]. However, another study found that cancer mortality costs in the European Union (EU) fell 11% in from €54.5 billion in 1995 to €48.6 billion in 2014 in absolute terms and 15% per-capita, mainly due to improved survival rates [9]. In the United States (US) productivity losses in 2020 from cancer mortality were estimated using modelling at US$147.6 billion. When imputed earnings lost due to caregiving and household activity were included, the projection rose to $308 billion [68], while a study in Iran found 1,112,680 years of productive life lost (YPLL) in

2015, for the Iranian population of approximately 80 million people, due to premature cancer-related deaths [69].

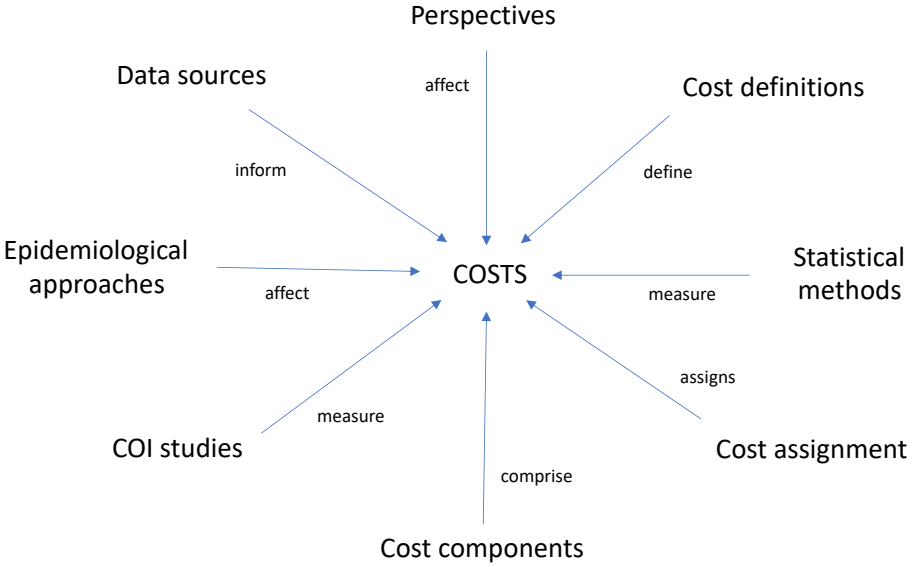
2.2.4 Long-Term Survival

Cancer patients who survive often suffer significant enduring health effects of the disease and its treatment, such as infertility, sexual dysfunction, cognitive impairment, and premature ageing [70]. Survivors can also have a higher risk of developing second cancers, which can incur further financial burdens even decades after diagnosis [71], while financial distress and hardship are common, particularly in the US [72]. Ongoing surveillance may be required, adding additional stress to patients and extra healthcare resources. Costs can increase with survival as patients need continuing courses of treatment and need additional healthcare due to the presence of comorbidities [6]. The long-term effects of cancer on employment are not well known, but some evidence suggests employment is affected up to 10 years after the diagnosis [73]. Long-term effects on employment can arise from lasting physical or cognitive impairment or from hysteresis, where individuals suffer "scarring" from previous periods out of employment. Losses due to productivity are not limited to the workplace and may adversely affect normal functioning, possibly causing a need for help with daily tasks and raising the demand for professional care services.

2.3 Measurement of Costs

A diagram of the topics presented in this section is shown in Figure 2.1.

Figure 2.1.: Visualisation of topics related to measurement of costs



2.3.1 Defining Costs

Before examining how costs are measured in the literature it is worth considering what is meant by *costs*, as the meaning can vary subtly in different contexts, and is often used interchangeably with other terms such as *spending*, *expenditure*, *losses*, *outlays*, *resource use* and *burdens*. In some contexts the interchange of terms may matter little, while in others the substitution of one for another might imply different prescriptions. For example, if told that *spending* on cancer is disproportionately high, a policymaker might be tempted to decrease expenditure on cancer care. Another policymaker, if told that the *costs* of cancer are disproportionately high, might be tempted to increase expenditure on cancer care—if the costs in this case are simply existing expenditure then the existing misallocation of resources is compounded [74]. In addition, economists and accountants have different definitions of costs. Economic costs are defined as lost opportunities, while accounting costs are actual outlays on resources [75]. In the context of healthcare, Hodgson (1982) defined costs as

> ...the value of resources used, resulting in forgone alternatives, and resources lost due to morbidity and mortality. [76]

Whether economic or accounting, costs are typically quantified in monetary units such as dollars and pounds sterling. While this may seem unremarkable, it is not without problems, as many costs do not represent actual purchases, and the validity of assigning monetary values to healthcare resources is not universally accepted; the difficulties of standardising monetary costs across studies have prompted a recommendation to openly consider and discuss whether and why the inclusion of a monetary outcome is appropriate [77].

2.3.2 Cost-of-Illness Studies

While economic evaluations such as cost-benefit and cost-effectiveness analyses generally include costs, the standard non-analytical method for describing the costs of a disease is the cost-of-illness (COI) study. COI studies are widely used to advocate for public health, influence and prioritise policy [78, 76], and distribute the burden of illness to society [76]. COI studies are descriptive rather than analytic, do not test a specific hypothesis, and have been credited with providing information that enables cross-country and other types of comparison [79]. However, it has been claimed that they suffer from discrepancies in methodologies and reporting that limit their utility [80] and that inherent theoretical limitations lead to bias and circularity [74]. Despite controversy over their methods and ultimate value, COI studies are widely produced,

and generally focus on the costs of a particular disease, or groups of diseases, for aggregated individuals who suffer the disease.

2.3.3 Cost Components

To facilitate measurement and reporting, the costs of illness can be broken down into components such as those shown in Figure 2.2. Costs are often classified into three broad categories: direct costs, indirect costs and intangibles. Direct costs are those for which a good or service is procured or received, which may or may not be healthcare. Examples include primary care, outpatient costs, inpatient costs, social care, medicines, ongoing prescriptions and out-of-pocket costs. Indirect costs are those that occur due to losses in productivity, where no actual good or service is received or procured. Indirect costs can also refer to other types of cost, hence it has been suggested that the term *productivity losses* be used to avoid confusion [81]. Intangible costs are those that cannot be easily assigned a monetary burden, such as suffering, changes in social status and disruption to family routine [82]. Government payments such as incapacity benefits and other social security payments are typically not counted, because these are considered transfers of resources from one party to another rather than irretrievable losses, thus their inclusion would incur double counting [76]. Returning to Hodgson's definition of costs [76], direct costs relate to *resources used*, which are opportunity costs that generally include a decision on how to allocate resources (although in reality choice may be limited). Indirect and intangible costs relate to *resources lost*, which generally do not involve decisions on resource allocation and therefore cannot be strictly regarded as opportunity costs. A review of COI studies found that the five components most commonly included were hospital costs, outpatient costs, drug costs, productivity losses and laboratory costs [83]. As will be discussed below, the distinct natures of different categories of cost components can make their combination problematic and complicate the comparison of studies.

Figure 2.2.: Components contributing to the cost of a disease

DIRECT COSTS		**INDIRECT COSTS**	**INTANGIBLES**
MEDICAL	NON-MEDICAL	PRODUCTIVITY	NON-FINANCIAL
Medications	Transport & parking	Absenteeism	Suffering
GP visits	Childcare	Presenteeism	Social status
Diagnostics	Heating	Mortality	Life routine
Inpatient stays	Baby-sitting	Early retirement	Relationships
Outpatient visits	Home changes	Loss of leisure	Self-perceptions
Surgery	Relocation	Everyday tasks	
Radiotherapy	Clothing		
Rehabilitation	Dietary changes		
Social care			
Home care			

Productivity Losses

Productivity losses can comprise a substantial portion of overall costs, but their inclusion is not universal, and guidelines from the National Institute for Health and Care Excellence (NICE) in the UK recommend their exclusion from economic evaluations [84]. While excluding productivity costs can underestimate the overall cost of a disease, there are numerous theoretical and practical problems that make inclusion problematic. Some economists argue that in cost-utility evaluations the losses accruing to lost productivity are already included in the quality-adjusted life year (QALY), hence including them on the costs side would incur double counting [76]. Suitable data on productivity losses may be difficult to obtain and interpret. Furthermore, the monetisation of productivity losses has been noted as a major difficulty, with a lack of established and validated methods [85]. Even within studies that include productivity costs there are differences in the methodology to estimate the costs. Two commonly used methods are the human capital method (HCM) and the friction cost method (FCM), each of which has advantages and disadvantages.

The HCM tends to predominate in COI studies [3]. It is based on marginal

productivity theory, which attempts to explain the determinants of wages through the marginal productivity of workers [86]. Costs are derived from the expected wage that would have been earned if the patient had continued working, and are usually calculated from the median wage in the population under study [87]. It is assumed that the reduced productivity due to the disease cannot easily be replaced. However, the underlying theory of marginal productivity struggles to explain labour market imperfections such as discrimination [76, 86]. Additionally, marginal productivity cannot hold in all industries as many do not have marginal products [88]. Methods to adjust for these issues have been proposed [86], however, a further limitation of the HCM is its assumption of irreplaceable labour, which will lead to overestimations of the costs to society [89]. The FCM, an alternative approach to estimating productivity costs, was devised to address these issues [87].

The FCM takes the assumption that a pool of replacement labour exists in a workforce, hence if a worker's productivity drops due to disease it can be replaced, with only the cost of replacement being large enough to be accounted for. This cost is known as the friction cost [90]. Productivity costs derived from the FCM tend to be lower than those derived from the HCM, and, as no agreed reference standard exists, it is not known whether the lower costs are more accurate [91]. The FCM is a more complex method than the HCM and requires knowledge of market conditions that limit the method's practicality [91]. The assumption of a pool of easily replaceable labour may not hold in all labour markets, and the method is sensitive to labour market conditions, hence limiting the external validity of studies based on this method [91]. An additional estimation method is willingness-to-pay (WTP), which elicits individuals' preferences for reduction of illness or mortality by such methods as surveys, discrete choice experiments and wage differentials for occupations of varying risk [89]. Barriers to implementation include the difficulty of obtaining data on individuals' preferences [89].

In addition to theoretical divisions over the choice of estimation method, there are practical difficulties with measurement. Productivity costs may comprise multiple sub-components, and their inclusion may be limited by the availability of data. Sub-components include losses due to mortality, early retirement, unemployment, absenteeism, presenteeism and loss of leisure time, where absenteeism refers to time taken off work and presenteeism relates to loss of productivity at work. Practical difficulties exist in measuring and costing these components, which are likely to be more pronounced for presenteeism than for absenteeism [85], hence a multi-method approach has been suggested to account for differing job types [92]. Despite the

practical and methodological difficulties in measuring productivity losses and assigning them a monetary cost, ignoring them is likely to underestimate the cost of a disease to society, possibly substantially [79, 76]. Productivity losses for all diseases in the US were estimated at US$260 billion in 2007 [85], while a UK study found that 33% of employed cancer patients stopped working after their diagnosis [93]. Downward bias for poorer nations could also occur if the HCM is used to estimate productivity costs, because lower average wages will result in lower estimates of monetary losses [76]. A suggested solution to these problems is to quantify indirect costs in non-monetary terms that are appropriate to the particular study, for example, standard epidemiological measures such as relative risk (RR) for binary outcomes or economic measures such as YPLL [76].

2.3.4 Study Perspective

When measuring the costs of a disease, the choice of perspective, i.e., the party to which the costs accrue, will determine which components are relevant and consequently the overall cost. In COI studies the perspective is commonly chosen from those of the patient, healthcare provider, government, employer or society [89]. The perspective of the healthcare provider tends to be useful in economic evaluations where the efficiency of alternative interventions is being evaluated [94]. The patient's perspective is less commonly reported, with around one in ten studies taking this perspective [95], and may be more useful for policy-making that focuses on support to individuals. It has been argued that the societal perspective should be the default one, as health economics has foundations in welfare economics, which is concerned with society's welfare, and consideration of costs from the health payer's perspective may not maximise welfare because of costs to other parties such as social services and the families of patients [94].

2.3.5 Epidemiological Approaches

Epidemiological approaches to COI measurement are related to the epidemiological concepts of incidence and prevalence. Prevalence costs are those accruing to all individuals with a disease in a population during a defined period, e.g. one year. Incidence costs are those accruing to individuals from a defined start time, usually the cancer diagnosis, over a set period [96]. A variation of the incidence approach is the phase-of-care (POC) approach. This approach breaks the cancer-care trajectory into distinct phases relative to the cancer diagnosis: initial (treatment), continuing (post-treatment) and terminal (end-of-life). Costs are then calculated over each phase

[96]. Each approach has strengths and weaknesses, and the appropriate method should relate to the study objectives [96]. Prevalence costs are useful for describing the total costs of a disease to society, but may not be sufficiently detailed for other applications such as regression modelling or sub-group analysis [89]. Incidence costs can impart the dynamics of individuals and populations, but obtaining the necessary data may be a problem, and the overall magnitude of costs will be determined by the choice of time horizon [89]. The phase-of-care approach highlights expenditure in periods of interest and can make better use of data with low counts, which may be important for less common cancers [21], however incorporating discounting, as discussed next, may be problematic [97].

2.3.6 Time Horizons

Discounting

If the costs of a disease accumulate over several years, the question of how much weight to assign to future costs arises. Standard economic practice when dealing with future costs and benefits is to assign progressively lower values as one looks further into the future, based on the model of discounted utility (DU). This model reduces the motives for temporal choices to a single parameter known as the discount rate [98]. Motives include time preferences and beliefs about the future, hence the discount rate incorporates uncertainty (though it may not be explicitly considered). The model is accepted by economists as an accurate representation of observed behaviour and as a prescriptive standard for public policy [98]. However, the assumptions underpinning the model are shown to be violated in observed behaviour. For example, the magnitude effect describes how individuals discount small outcomes more than large ones [98] while hyperbolic discounting describes a decreasing discount rate over time and has been observed in health-related behaviour [99]. Losses and gains are not discounted equally and behaviour such as delayed satisfaction and bringing losses forward have been observed [98]. Time preferences have been used to explain addictive and disease-promoting behaviour [99], and even to attribute rationality (as in utility maximisation) to such behaviour [100], while alternative theories include preference reversal [101] and habit formation [98]. It is argued that the discount rate should be based on the time preference rate for health and the same rate used for costs [102]. However, it is not clear how the transition from a positivist description of individuals' behaviour to a normative prescription of public policy is derived. Furthermore, individuals may undervalue future health states or later reverse their preferences [101]. There are also ethical issues around intergenerational fairness. The

normative aspects of discounting schemes have been denounced as ethically flawed due to disadvantaging future generations (particularly in relation to climate change after controversy over the Stern Review's choice of a lower than expected discount rate [103]), prompting the development of new methods to address such issues [104]. Discounting has also been shown to introduce bias and complexity, because discounted costs show the present value of expected costs, not the expected costs themselves, which may give a misleading impression of actual future expenditure [97]. Discounting aids evaluation of competing investments with future costs and benefits, as the investments represent opportunity costs over other investments such as a simple cash-holding investment with interest [98]. Consequently, discounting may be more suitable in evaluations of competing interventions than descriptive studies of realised costs, although issues of intergenerational fairness may persist if choices seem attractive on the basis of passing costs to future payers [98]. A difficulty for COI studies is that their results may be used for either purpose, causing a dilemma over whether to present discounted or undiscounted costs. One solution is to present both [21]. An alternative, where incidence costs are measured over several years, is to present undiscounted costs for each year of interest, thus allowing other researchers to discount at whatever rate is appropriate [105, 81].

Censoring

Considering the high mortality rates of cancer, the question of how to incorporate the costs, or lack of them, for patients who die would seem of importance. However, there is little standardisation in study methodologies around censoring, although the approach may substantially influence costs [77]. An approach seen in incidence studies is to censor patients at the end of each study period, so costs only accrue to those patients alive at the start of each period [97, 105]. This approach is useful for highlighting the costs to patients with the disease, but can overestimate the overall costs to society in at least two ways: the first is that end-of-life costs, which are known to be high regardless of cancer diagnosis [106], are included for those who die during the study, but not for those who survive the study period; therefore, these costs are deferred rather than zero; the second is that it does not account for costs that patients with the disease would have accrued had they survived, assuming the disease to be the cause of death. The phase-of-care approach avoids the first problem by including the end-of-life phase for survivors and for those who die (including controls in a comparative study) [21]. An uncensored approach of assigning deceased individuals zero costs throughout the study time horizon avoids the second problem, but can be uninformative for diseases with poor survival such as lung cancer. The

appropriate choice will depend on the goals of the particular study, however methodological variation across studies may reduce comparability [77].

2.3.7 Cost Assignment and Attribution

While medicines can be readily monetised, these comprise only a portion of healthcare costs. Additional costs include the labour of healthcare professionals, equipment, buildings, utilities and other factors [107]. Within public healthcare systems there are seldom authoritative accounts of expenditure that can be consulted when estimating the costs of an illness, consequently, costs must be estimated indirectly from records of healthcare use, such as counts of hospital episodes [108]. The assignment of reference costs to routinely available data from hospital episode statistics is an accepted primary method of healthcare costing, but cannot be considered a "gold standard" [109]. An outstanding problem is how costs can be adequately assigned to units of resource use, as definite reference costs are usually lacking [108]. Various methods have been devised to tackle this issue, including assigning costs per-diem, per-episode, per-stay, and using healthcare resource groups (HRGs)—also called diagnosis related groups (DRGs). Each method has advantages and disadvantages, and the most suitable choice will depend on the goals of the particular study [108]. Analysis using Scottish hospital admissions data found considerable variation in final costs estimated by each method [108]. Costings using HRGs were found to be the most accurate but are dependent on high-quality reference costs for the health groupings appropriate to the health system under study. The Scottish tariffs used by Geue et al. (2011) [108] were developed using English reference costs, and rely on the assumption of resource differential equality which may be unrealistic. Per-diem costing assumes that all days spent in hospital during a single stay take the same cost value, and hence does not differentiate between treatment and recuperation periods during a stay, which are likely to have different daily costs. Per-diem costs are therefore believed to have an upward bias, on account of longer stays being overestimated [108]. An alternative is per-episode costing, the most simple method of which does not take length-of-stay into account and merely derives costs from national averages of episode costs. A more complex method splits episodes into fixed and variable costs, the latter incorporating costs related to length-of-stay. As this method can, to some extent, distinguish treatment from recuperation periods it is less prone to the upward bias found in per-diem costing [108] while still accounting for longer stays. Substantial differences were found between methods with HRG costings producing the lowest estimates and per-diem the highest [108]. In the absence of a "gold standard" it is not clear which

method produces the most accurate costings. The expected bias in per-diem costing may make other methods more attractive unless the length-of-stay is related to the study question. If whole-of-population data are available, such as those from hospital episode statistics, the average episode costs should, by the central limit theorem, approach the true average, making per-episode costing attractive due to its simplicity, though it may be less accurate for sub-groups. New costing methods may consign this issue to a historical one; the Patient Level Costing System (PLICS) records patient-level costs for each episode of healthcare and may provide a "gold standard" for future studies [110].

At a broader level, cost measurement may take either a bottom-up or top-down approach. Top-down costing, also known as the epidemiological approach, takes aggregated population-level data and divides it by the size of the population to estimate costs for an average person. This method is suited to the prevalence approach [96]. Population-attributable fractions (PAFs) may be used to estimate the costs attributable to the disease under study [89]. Bottom-up costing, which is more appropriate to the incidence approach, uses individual-level data to estimate costs for each patient. The resulting patient-level costs may be averaged or combined with population-level statistics to estimate the total costs of a disease. An additional method is the econometric approach, which matches an exposed and control cohort then estimates costs using either mean differences or regression modelling [89]. Costs are commonly reported in COI studies as total, excess (also known as net) or condition-attributable. Total costs are simply the sum of all costs measured for a particular patient. Excess costs are equal to the mean total costs for those with the disease minus the mean total costs for similar individuals without the disease [96]. Excess costs are particularly valuable for examining how costs change over time, as changes in total costs may simply reflect the effects of ageing but a control group will be required [111]. Condition-attributable costs are costs specific to the condition under study [96]. As correct attribution of costs to a particular disease is a non-trivial problem, this method may require more complex methods but has the advantage of eliminating the need for a control group [96].

2.3.8 Data Sources

While prospective studies can be used to measure costs, many COI studies now use retrospective data derived from routinely collected administrative datasets, which have been used to measure cancer costs in many countries [112, 113]. Such datasets may be held by private insurance companies, public insurance bodies, charities, public

healthcare providers, social security bodies and other public organisations. The pre-existence of routine administrative datasets eliminates the burden of primary data collection, while other advantages to researchers are large samples (often covering entire populations), lower attrition, and fewer measurement errors than survey data [114]. Other claimed advantages include discreet data collection that allows rigorous analysis of many economic variables not previously available, and low marginal cost [115]. Novel quasi-experimental research designs are possible and in some cases it may be feasible to link records to existing experiments and follow outcomes over time [116]. It has been claimed that the efficient use of large administrative datasets can reduce budget deficits [117] and also drive evidence-based care practice and policy [66]. Further claims are that patient-level administrative data provide new opportunities to understand costs for cancer patients, and are more cost-effective, with a broader scope, than primary data [66].

Administrative datasets are not without issues, however. Some may lack data on comorbidities [118], while productivity costs and other non-medical costs tend not to be available in health registries, making the inclusion of these components less frequent in studies using such data [79]. The large size of administrative datasets can incur substantial computational burdens, prompting the development of approaches to process massive data more efficiently in COI studies [119]. Bias may not be eliminated and it has been claimed that, in addition to biases found in other studies, routinely collected health data can add additional ones due to problems with coding, missing variables and other missing data, and changes in eligibility over time [120]. Additionally, there may be little documentation to guide researchers and the data may not be suitable for the research question at hand [115]. Reproducibility may be a problem and political will at national level may be necessary to develop capacity for data storage and analysis capacity [121]. With an ever-increasing number of studies using routinely collected health data, reporting standards have been developed, such as the Reporting of studies Conducted using Observational Routinely-Collected Health Data (RECORD) guidelines, to promote transparency of reporting [120].

The use of public administrative datasets raises further questions. For instance, how can informative data be published while ensuring the individuals under investigation are not identifiable? Anonymisation of data can ameliorate privacy issues, but this may be insufficient to reassure populations. While de-identification techniques are available, a study using generative modelling has cast doubt over their effectiveness [122]. There is also security of data to consider. Given the prevalence of cyber attacks, can administrative data be kept secure? And does the use of such data for research

purposes increase the risk of data breaches? In addition to such questions are ethical ones. For instance, is it ethical to use data that people have not consented to? If not, how could such consent be obtained? For many datasets, achieving consent may be difficult or impossible [115], If explicit consent is not deemed necessary, should people be allowed to opt out, and, if so, how could such an optout be obtained and implemented? The controversy surrounding NHS England's Care.data project suggests that data sharing may be unpopular if data subjects are not well informed and if they suspect that private parties will profit [123]. Despite the abandonment of this project, NHS England, through its data body NHS Digital, is again attempting to share patient data with an external organisation resulting in more controversy over patients' rights and threats of legal action [124, 125]. This raises the further issue of the legality of using administrative for purposes other than the original ones.

To facilitate research and collaboration, while minimising privacy violations trusted research environments (TREs) have been created [126]. They are intended to gain the trust of the public and patients, data custodians and researchers [126], however it is has been suggested that these environments do not operate on trust, but rather the lack of it [127]. This viewpoint has been contested [128]. Examples include OpenSAFELY in England [129], the DataLoch in Scotland [130], the European Open Science Cloud (EOSC) [131] and the Australian Research Data Commons (ARDC) [132]. The global Covid-19 crisis has accelerated the dissemination of large amounts of data, with OpenSafely reported to have delivered analysis of 58 million pseudonymised patient records in five weeks from the project start [129]. The data sources discussed are summarised in Table 2.1.

Table 2.1.: Summary of data sources

Name of data store	**Location**
OpenSAFELY [129]	England
DataLoch [130]	Scotland
European Open Science Cloud (EOSC) [131]	Europe
Australian Research Data Commons (ARDC) [132]	Australia

2.3.9 Statistical Methods

Health costs tend to have non-normal distributions with positive skew and heavy tails, making analysis challenging [133]. Multimodality is often present, and because many patients use no healthcare it is common to have distributions with high mass at zero. Non-parametric statistical methods can deal with such data, but are typically

neglected due to greater interest in the population mean as opposed to rank order statistics [133].

Austin et al. (2003) list three main reasons to model cost data [134]:

1. to understand the factors influencing costs
2. to determine the economic burden of disease
3. to evaluate the relationships between constructs such as socioeconomic status and care delivery costs.

While several types of regression are possible and relatively unbiased, generalised linear models (GLMs) have been found to provide the highest predictive accuracy however the choice of model should depend on its appropriateness for the particular data set [134] and simple models are preferred for reasonable sample sizes where sample means approach normality [133]. For counts of healthcare episodes and other units of resource use, Poisson and negative binomial models are suitable, whereas for data with substantial zero counts, two-part models may be more appropriate [134]. Linear models are usually inappropriate because cost data do not normally take negative values but may be sufficient to identify factors driving costs where the magnitude of the coefficients is of lesser importance [134]. Future research may incorporate mixed models and Bayesian approaches that incorporate distributional shapes in the priors; however, until analytical frameworks that deal with complex methods are developed, complex methods may bring few advantages over simpler ones [133]. Of note is the Kaplan-Meier sample average (KMSA) method, which can be used to model total cost estimates by combining phase-specific monthly costs with monthly survival probability estimates [135]. An advantage claimed for this approach is that it can account for the censored nature of the available data [97]. Although widely used in health evaluations [136] simulated data are less commonly seen in COI studies, but possible simulation methods include discrete event simulations [137] and Markov models [138, 139] though other modelling techniques may also be feasible [140, 141]. Table 2.2 summarises these methods.

Table 2.2.: Summary of statistical methods used in measurement of costs

Statistical method	Uses	Strengths	Limitations
Generalised linear models	Estimating costs and resource use, determining risk factors, confounder adjustment	Deal with zero counts, no outcome transfornation needed	Interpretation of coefficients may be less straightforward than OLS
Poisson regression	Estimating costs and resource use, determining risk factors, confounder adjustment	Deals with count data such as numbers of healthcare episodes.	Different types of episodes difficult to aggregate
Negative binomial regression	Estimating costs and resource use, determining risk factors, confounder adjustment	Deals with count data such as numbers of healthcare episodes	Different types of episodes difficult to aggregate
Linear regression	Estimating costs and resource use, determining risk factors, confounder adjustment	Coefficients are simple to interpret	Normality, negative values, problems with high zero mass
Two-part models	Estimating costs and resource use, determining risk factors, confounder adjustment	Deals well with high zero mass, well suited to healthcare costs	Model is in two-parts so coefficients are more complex to present and interpret
Kaplan-Meier sample average	Estimating incidence costs and resource use across multiple phases of care	Can estimate longer-term costs from data over short time frames	Survival probability at each individual phase needed
Markov models	Simulating data, estimating costs and resource use with transitions present	Deals with discrete transitions and decisions	Many branches can add complexity
Discrete event simulations	Simulating data, estimating costs and resource use with discrete events present	Deals with discrete events and decisions	Many events can add complexity

2.3.10 Measurement Issues and Heterogeneity of Results

COI studies are used to advocate for public health and influence policy, however discrepancies in methodologies and reporting may limit their external validity and ultimate utility [80]. It has been claimed that theoretical limitations lead to bias and circularity in estimates [74] and that a lack of transparency in reporting can hinder interpretation [80]. Variability in terminology and other aspects of reporting makes synthesis challenging, with considerable heterogeneity in the results reported between studies [66, 77, 142, 66]. While broad cross-national patterns are observed, there is considerable variation between countries, which limits comparability [7, 143], and even within countries high variation is observed [77]. Isolated communities tend to incur higher healthcare costs as the cost of building and maintaining infrastructure in remote areas is higher while the availability of trained medical personnel is lower, and the higher costs of travel may cause very ill patients to forgo treatment in order to die at home with loved ones [82]. International differences that can affect cancer costs include differing demographics and variations in survival rates [7]. Another source of variation is the price of drugs, which can vary substantially between countries [143]. In Europe, health authorities are often able to negotiate prices for cancer drugs with drug manufacturers, while in the US, two of the largest payers of cancer drugs, Medicare and Medicaid, are required to offer most approved cancer drugs and must pay the manufacturer's full price for them [144]. A consequence is that private insurers have little capacity to negotiate price, leading to cancer drugs costing twice as much on average in the US compared to Europe [44], and for some drugs the discrepancy is considerably greater [144]. Per-capita spending in the US on all healthcare is more

than twice that in the UK [144], and a significant influence on this discrepancy is drug costs. In addition to drug prices, there are major differences in health systems that influence costs sufficiently to limit the comparability of COI studies. Results may be bound to the particular nation under study unless steps are taken to standardise the methods and data [143]. An additional source of heterogeneity is the treatment of indirect costs. Cancer drugs in the UK must meet the NICE cost-effectiveness guidelines that specify the exclusion of productivity losses in health technology evaluations [84].

Identifying and synthesising studies measuring cancer costs has been noted as significantly challenging [66]. A systematic review of prostate cancer showed considerable variation between studies even for populations within the same country and with comparable time horizons [145], while a systematic review of melanoma found a similar pattern of high variation between studies [142]. Variability in reporting and terminology makes the high number of potentially relevant abstracts unrealistic for researchers to review [66]. Despite the heterogeneity of costs there are common patterns across studies. The cost of cancer is reported to be substantial [96]. Costs vary substantially by cancer site, and this is seen within studies as well as between studies [105, 135, 146, 21, 63]. The association of a later stage with higher costs may result from higher mortality in cancers such as lung cancer [21, 135], however the costs of prostate cancer are generally reported as lower than those for lung [105, 135], despite detection being common at an advanced stage. An explanation may lie in prostate cancer's lower mortality [145], however, results for breast cancer and colorectal cancer show comparable costs to those of lung cancer [63, 21, 105], despite longer survival, suggesting that the association between costs and survival is not a simple one.

2.4 Summary and Discussion

Section 1.1 described how trends in cancer risk factors, treatment, and survival are making cancer a longer-term condition. The overall effect on costs to healthcare systems and wider society will depend on a number of factors. The trend in total oncology spending suggests increasing costs [7, 6], but much of the existing increases are due to new, expensive treatments, which are borne largely by the US healthcare system rather than high-income countries in Europe, which have more power over drug pricing [44]. Improvements in screening, treatment and surveillance could improve efficiency and lower costs [7], but longer survivorship will require additional surveillance

and healthcare [7, 6], particularly where survivors have comorbidities [6]. Other new technologies such as telemedicine, smart gadgets and robotics may reduce hospital visits and consequent costs [8]. The decline in tobacco-related cancers may also impact costs, and public health initiatives to reduce tobacco use might be emulated for other risk factors. Beyond healthcare, raising the pension eligibility age implies that, *ceteris paribus*, more working-age individuals will have cancer, which could place additional burdens on employers, individuals and social security systems.

2.4.1 Measurement

Considering the heterogeneity seen in many aspects of COI studies, care must be taken when interpreting and comparing their results. There is a lack of clarity and transparency around what is meant by *costs*, and the distinction between unavoidable losses and chosen expenditure is seldom made explicit, which may compound inefficiencies. As noted by Shiell et al. (1987), if policymakers dedicate resources to disease based on previous misallocation, the initial mistakes will be compounded [74]. However, despite their limitations, COI studies can be useful to policymakers, healthcare professionals and economists in deciding how to allocate scarce resources [89]. Existing approaches can capture long-term costs and have been used for long-term chronic diseases such as diabetes [147]. For cancer, the phase-of-care approach may be attractive where long-term data is lacking [97], while incidence costs measured with long-term data can capture cohort dynamics the phase-of-care approach may miss [96]. As cancer survival improves and wider social impacts, such as on pensions and other social services, become more apparent, the case for the societal perspective grows stronger.

2.4.2 Productivity Costs

The debate over the inclusion of productivity costs is unlikely to be resolved quickly. Costs arise from choices around how to commit resources, not from disease directly, which in most cases is not a conscious choice [74]. As losses from productivity are not the result of decisions involving resources they cannot be considered opportunity costs. However, this strict economic interpretation of costs, while perhaps more useful for evaluating the benefits of one or more interventions, could underestimate the total impact of a disease to society, particularly in poorer nations where populations may suffer highly impaired productivity due to lack of healthcare access [76]. In high-income nations, nominally higher wages and stronger currencies are likely to increase the

monetary magnitude of productivity costs. As described in Section 2.3.10, difficulties in measurement together with theoretical disagreements make the inclusion of such costs in COI studies problematic, however to ignore these costs could misrepresent the burden of disease and give misleading comparisons with other diseases [79, 76]. Hence the separate reporting of productivity losses in non-monetary units such as YPLL should be considered [76]. While obtaining data on productivity costs is not straightforward, the possibility of linking administrative public datasets with routine healthcare data presents an opportunity for gaining insights into the long-term effects on employment and earnings. Longitudinal surveys, with their rich socioeconomic data over time, may also be of value, however health variables may here be lacking. Hence, building a complete picture of disease costs may require synthesis of evidence from multiple data sources and reporting of outcomes in multiple dimensions.

2.4.3 Unit Cost Assignment

The problem of assigning costs to units of healthcare use may be resolved if PLICS becomes more widely adopted. Meanwhile, costing using HRGs seems likely to remain the most precise costing method and the most appropriate for health evaluations where sufficient reference data exist. However, given the lack of methodological unity across COI studies and the notable heterogeneity of results, the gains in precision may not justify the increased complexity over simpler methods such as per-episode costing [108]. Furthermore, if the HRGs used are derived from a different health system from that under investigation, bias may be introduced [108].

2.4.4 Discounting

While discounting may be appropriate for comparing the present value of competing investments or healthcare interventions, it is harder to justify its applicability to non-evaluative descriptions of costs in the past [97]. The existence of public health initiatives to reduce harmful long-term behaviours like tobacco smoking, could suggest that public health organisations set a different discount rate on future healthcare costs than individuals, or make decisions about future costs using alternative criteria [98]. Where costs are merely descriptive of past or present expenditure, discounting is unnecessary [81]. Where costs are a forecast of future costs, the case for discounting is stronger given that the rationale for discounting is made clear. Other approaches are possible. For example, where incidence costs are calculated over a number of periods, the presentation of undiscounted costs during each period could allow other

researchers to set custom discount rates and perform sensitivity analysis for a range of discount rates.

2.4.5 Data

As described in Section 2.1, the accumulation of routine healthcare data and other administrative data presents opportunities for studying healthcare costs and wider societal costs across entire populations over long time frames. Connecting different sets of data could provide valuable insights into cost drivers for sub-populations. However, the aggregation of large datasets may not solve the inherent weaknesses in COI methods, furthermore variables of interest may be missing. As described in Section 2.3.8 there are issues around privacy, ethics, legality, and security that could make data access problematic, but which have spurred initiatives to enable faster access to anonymised datasets and to link health data with longitudinal survey data. Future researchers may have access to a wider range of cost components bringing them closer to measuring the full costs of cancer to society.

2.4.6 Opportunity Costs

While great progress has been made in detecting and treating cancer, it should be kept in mind that a major goal of much cancer research is the desire to earn profits [45], while other spending, such as on prevention or on treating other diseases, may be more efficient at maximising health outcomes for the same expenditure [23]. While extending lifespans is a worthy goal, there may be additional societal costs such as increased demand for healthcare, social care, pensions and housing that will strain public finances while being ignored by healthcare evaluations.

2.5 Conclusions

Cancer is becoming more of a long-term chronic disease due to improving diagnostics and treatments. Understanding the effects on healthcare, employment, social care and other social services is necessary to efficiently provide these services. Theoretical issues around productivity losses make their incorporation into a single monetary unit problematic, however as cancer becomes more long-term and retirement ages rise, it is necessary to better understand the long-term effects of cancer on employment and other aspects of productivity. Growing repositories of administration data offer opportunities to investigate the long-term costs of cancer. Linkage of such datasets in

Scotland could provide information on healthcare use and employment, which could enhance understanding of the wider costs of cancer over the longer term. In the chapters to follow, I will describe how I used such data to answer these questions and the challenges involved. The next chapter will describe how I used routine healthcare data from NHS Scotland to measure the costs of healthcare for people with a cancer diagnosis.

3 Eight-Year Healthcare Resource Use of Cancer Patients in Scotland Using Linked NHS Datasets

3.1 Introduction

3.1.1 Background and Rationale

Section 1.1, described how cancer is becoming more of a chronic long-term disease, due to changes in risk factors, screening and treatment. The likely effects on healthcare use and associated costs are uncertain. Improved detection in conjunction with more efficient treatments may reduce costs, but new treatments and other technologies may be more expensive than existing ones. Improved survival may increase healthcare use due to ongoing surveillance, recurrences of cancer [6] and the presence of comorbidities [148]. Additionally, there may be long-term effects in survivors such as cognitive impairment, sexual dysfunction, infertility and premature ageing [70]. Chapter 2 examined how the costs of cancer can be measured. A possibility lies in the use of linking administrative healthcare data to measure the costs of resource use. In this Chapter, I will describe how I used linked data from NHS Scotland to measure the resource use and associated costs of people with cancer in Scotland following their diagnosis.

To understand how a cancer diagnosis impacts healthcare use and associated costs I carried out a literature search on the resource use and costs of cancer in Medline, Embase, Google Scholar, Web of Science and Cochrane Reviews. I first identified studies carried out in the Scottish population. Systematic reviews detailing other populations were identified, however, these described highly heterogenous results for particular cancers and were of limited value. I further identified studies that examined patterns of cancer use across cancer types and across time. Other studies were identified where they highlighted a particular aspect of costs. The studies are summarized in Table 3.1. The Ovid search strategy is presented in Appendix A.

Table 3.1.: Summary of literature results on the costs of cancer

Study	Year	Cost base	Population	Methods	Notes	Cancers studied	Results summary (2018 PPP)
Hall et al.	2015	2012 sterling	NHS England Trust	15 month cumulative hospital costs	Uses routinely collected NHS data	Breast, prostate, colorectal	Breast: £14,281; prostate: £4,220; colorectal: £14,336
Laudicella et al.	2016	2010 sterling	NHS England population	9 year incidence + 5 year prevalence	9 year incidence (total costs of care for patients alive at beginning of each year)	Lung, breast, prostate, colorectal	Lung: <65 £27,198, >=65 £24,975; breast: <65 £27,679, >=65 £27,323; colorectal: <65 £40,090, >=65 £39,932
Marti et al.	2015	2012 sterling	298 patients in ePOCS study England	Patient level data for 3 month period 12-15 months post-diagnosis	Reports healthcare, informal care and OOP costs separately and combined	Breast, prostate, colorectal	Breast: £380 / month; prostate: £134 / month; colorectal: £306 / month
Banegas et al.	2018	2015 USD	45,522 cancer and 314,887 controls using US health plan data	Total and net costs, 1 year and 5 year, using phase of care approach	Costs higher for <65. Costs presented by stage and <>65. Shown here are stage 3, 5 year costs for over 65s	Lung, breast, prostate, colorectal	Lung: £95,352; breast: £72,229; colorectal: £65,696
Yabroff et al.	2008	2004 USD	718,907 cancer and 1,623,651 non-cancer US Medicare patients (65+)	Net costs by phase of care using survival data to give 5 year costs	Surveillance, Epidemiology, and End Results (SEER) data linked to Medicare records. 5 year net discounted costs estimates for year 2004 shown	Lung, breast, prostate, colorectal, skin, bladder, non-Hodgkin lymphoma, head and neck, liver, oesophagus	Lung: M £33,977 F £35,503; breast: F £17,001; prostate: M £18,450; colorectal: M £34,868 F £24,105; skin: M £8,511 F £6,279; bladder: M £22,900 F £21,538; lymphoma: M £42,464 F £39,177; head and neck: M £29,156 F £39,177; liver: M £32,225 F £35,016; oesophagus: M £44,993 F £41,444

Notes: PPP = purchasing power parity, USD = United States dollar, OOP = out of pocket, M = male, F = female

The literature on the healthcare costs of cancer [1] indicates that the resource use and associated costs of cancer patients are substantial, but heterogeneity in research questions, methods and populations makes the synthesis of results challenging. For example, a systematic review of prostate cancer found costs ranging from 732 to 39,143 Canadian dollars, while in a systematic review for skin cancer the range of costs was 12,730 to 69,006 US dollars for stage III cancers. Differential exchange rates and inflation rates add to the complexity of comparing monetary costs across countries and time. To aid comparability, the costs that follow were converted to sterling using historical purchasing power parity (PPP) ratios from the Organisation for Economic Co-operation and Development (OECD) [149], then inflated to 2018 price levels using inflation rates from the Bank of England (BoE) [150]. While the magnitudes of costs show considerable heterogeneity across studies, patterns of healthcare costs across cancer types and time are believed to be more generalisable than specific results [151]. Two studies that measured total healthcare costs for people over 65 years old found mean per-patient costs of £24,975 versus £95,352 for lung cancer, £27,323 versus £77,229 for breast cancer, £28,209 versus £48,520 for prostate cancer and £39,932 versus £65,696 for colorectal cancer, in England and the US respectively [105, 135]. The discrepancies in cost levels are in spite of conversion to PPP and 2018 price levels, and the US study with higher costs having a five-year follow-up compared to a

[1] The literature search strategy is shown in Appendix A.

nine-year follow-up in Laudicella et al. (2016) [105]. Another study in England observed mean total per-patient costs of £14,281, £4,220, and £14,336 for breast cancer, prostate cancer and colorectal cancer respectively, again inflated to 2018 levels. The lower costs in this study compared to those in Laudicella et al. (2016) [105] were likely to have resulted from the shorter follow-up time of 15 months. In all studies prostate cancer had lower costs while breast and prostate were similar to each other in magnitude, though somewhat higher in Laudicella et al. (2016) [105]. This pattern was also seen in Marti et al. (2015) [146] with monthly per-patient total costs of £380, £134, £306 for breast, prostate and colorectal cancers respectively, again in the English population [146]. The higher costs in Banegas et al. (2018) [135], particularly for lung cancer, may partly have resulted from the phase-of-care costing method used, but are also likely to reflect the different healthcare systems of the UK and US. US studies commonly use either private insurance claims data, or Medicare data—which are limited to patients over 65 years old [21] while higher costs have been observed in patients under 65 years old [105, 135]. Spending on healthcare is considerably higher in the US than in Europe [144], with cancer drugs costing around twice as much on average [44]. Studies of cancer costs from other European countries may be more comparable to the UK, however the UK has poorer cancer outcomes [152] and its cancer spending of 5% of total healthcare spending is lower than the European average of 6% [153].

Looking at patient-level trajectories, the majority of healthcare use of cancer patients has been observed to occur in the year after diagnosis [63], however the relatively short time frame of this study may not have captured the cumulative effects of reduced survival and higher healthcare use in survivors over the longer term. Other studies found that healthcare use increased considerably as death approached [66, 105] even when this was up to five years beyond the diagnosis [105] and that the cumulative costs of cancer in subsequent years were comparable to, or even higher than those in the year after diagnosis [105]. Factors associated with high healthcare use include sex, marital status, ethnicity, geographical region [66], and stage [63, 135], while a small proportion of patients typically incur very high costs [146]. Considerable variation in costs has been observed between cancer sites, with prostate cancer found to have lower costs than breast cancer and colorectal cancer [63, 146, 135], and lung cancer [135].

Evidence on cancer resource use and associated costs in the Scottish population is limited, with studies focused mainly on subgroups of breast cancer using small samples and with limited information on costs. A study from 2008 found that cancer patients used more hospital care than heart failure patients with 20.5 bed days on average vs

20.2 [154]. The bed-use increased with age and deprivation increased the likelihood of frequent admissions. This study used linked hospital data but provided little information on costs. Another study found that cancers associated with tobacco smoking increased hospital use with longer stays and more admissions, but the sample was restricted to young women [155]. A 2007 study found that five-year treatment costs of breast cancer recurrences ranged from £10,000 to £37,000 for local recurrence and £14,500 to £20,000 for distant recurrence. This study only looked at healthcare related to recurrences rather than all healthcare [156]. A more recent online survey found that 64% of respondents with early-stage breast cancer used healthcare after their diagnosis [157]. The response rate of this survey was only 46% so the analysis may have suffered from non-response bias and surveys are known to suffer from other biases such as sampling bias [158].

The generalisability of other results to Scotland are complicated by the considerable differences in health within the UK, as described in Section 1.1.6. An additional problem when generalising results from England is that Scotland has a distinct healthcare system comprised of 14 regional National Health Service (NHS) boards. The high degree of centralisation across NHS Scotland offers opportunities to researchers, as routine healthcare datasets can be linked using a unique patient identifier number. The addition of an appropriate unit-costing method enables the aggregation of resources used in distinct NHS services, such as hospital stays and community prescriptions.

3.1.2 Aims and Objectives

While the costs of cancer have received considerable study, there are gaps in the literature around long-term costs and their magnitude in the Scottish population. This study aimed to increase understanding of the economic costs of cancer by measuring the long-term healthcare use of cancer patients in Scotland using linked administrative data. The study had the objectives below.

1. Describe the dynamics of healthcare costs and survival by charting trajectories.
2. Examine how costs vary between different cancer types.
3. Measure and compare risk factors for costs and survival.
4. Explore the potential of linking public-sector datasets to gain a wider picture of costs.

3.2 Methods

3.2.1 Study Overview

Study Design

The analysis took the form of a retrospective cohort study using patient-level data. Multiple routine health datasets were linked to measure the healthcare use of individuals over an eight-year period after a cancer diagnosis. A bottom-up micro-costing approach was taken, using reference costs from the Scottish Costs Book to estimate incidence costs over an eight-year period. Cost estimates were presented as cumulative eight-year costs, and were also stratified by year and by phase-of-care.

Setting

Scotland's public health system comprises 14 regional NHS Boards, covering Scotland's population of approximately 5.5 million people. This formed the study population. Eligible patients had a cancer diagnosis recorded in the Scottish NHS Cancer Register, known as Scottish Morbidity Record (SMR) 06 during the period 1 January 2009 to 31 December 2010. Each patient was followed for eight years ($8 \times 365.25 = 2922$ days) after diagnosis, with the last possible date of follow-up being 31 December 2018. Additionally, pre-diagnosis resource use was obtained for a period of five years ($5 \times 365.25 \simeq 1826$ days) before the date of diagnosis. The earliest possible date of the pre-diagnosis period was 1 January 2004, the last possible date 31 December 2010.

3.2.2 Participants

Eligibility Criteria

- The patient had an entry recorded in the Scottish Cancer Registry (SMR06).
- The date of diagnosis of the entry occurred during the recruitment period 2009–10.
- The patient was alive at diagnosis.
- The patient was at least 18 years old.

Exclusion Criteria

- The patient was under 18 years old at the time of diagnosis.

- The patient had a discordant death certificate (where the date of death preceded the cancer diagnosis).
- The patient's cancer was not a first cancer, i.e., a previous SMR06 record for the patient prior to the exposure window.
- The patient's cancer was identified during autopsy.
- The patient was not matched to any control in the linked datasets.

The study cohort was created in tandem with a matched cohort of similar individuals without cancer. The control cohort was not used in this analysis but the matching process affected the study numbers as a small portion of cancer records were unmatched. While it would have been ideal to include these records in the analysis, the low numbers would have caused disclosure issues around differencing of numbers between the analyses. As the proportion of unmatched individuals was <0.001% of all records, the effect on results was likely to be low. I selected the cohort by filtering the SMR06 dataset according to the above criteria. The full range of ICD10 (version 3) codes included in the cohort was C00–C96 excluding C44 [2]. I stratified the SMR06 cohort by the 10 most common cancers in Scotland, as defined by NHS Scotland incidence rates in 2019 with the ICD10 codes listed below. All SMR06 records belonging to other ICD10 codes were assigned the category *Other malignant neoplasms* and included in the analysis. This gave cancer groups as follows.

1. trachea, bronchus and lung (C33–C34)
2. breast (C50)
3. colorectal (C18–C20)
4. prostate (C61)
5. head and neck (C00–C14, C30–C32)
6. malignant melanoma of skin (C43)
7. kidney (C64–C65)
8. non-Hodgkin lymphoma (C82–C86)
9. oesophagus (C15)

[2] C44 is not recorded in SMR06 because non-melanoma skin cancers (NMSC) are excluded from analyses of all cancers to enable comparison with other countries that do not record NMSC data, and because only the first occurrence of the most common type of NMSC is collected due to high incidence [53]

10. bladder (C67)
11. other malignant neoplasms

Follow-up was achieved by linking SMR06 records with records in SMR00 (outpatient visits), SMR01 (inpatient episodes) and the Prescribing Information System (PIS) dataset, which records prescribed items dispensed in the community. Linkage was achieved via the Community Health Index (CHI) number, which is unique to each patient. The CHI number was replaced with an anonymised identifier during linkage. If a patient died or moved away during the follow-up period, measurement of resource use would continue at a rate of zero for each year after death. As a result, cost outcomes were recorded for all patients throughout the follow-up, regardless of whether they survived to the end of the eight years. For inpatient records, the time in days since the cancer diagnosis was calculated as the admission date minus the diagnosis date. For outpatient visits, the time since diagnosis was calculated by the outpatient visit date minus the diagnosis date. Due to very high numbers, individual records of prescribed items were aggregated over years, hence only the year of prescription was known rather than the exact dates. I calculated the number of years between the cancer diagnosis and the prescription by the prescription year minus the diagnosis year.

The study size balanced a number of factors. I required a sufficiently large sample to examine sub-groups over a follow-up period of several years, while recruiting from a period sufficiently recent that changes in treatment efficacy would not make results out of date. As PIS records were only available from 2009, increasing the study size while including PIS records would have meant reducing the post-diagnosis follow-up or increasing the complexity of cohort selection. Hence I used 2009 as the start period to allow the inclusion of prescription costs in the follow-up period. Although a large study size was desirable to increase statistical power, there were also computational demands to consider, as well as issues around disclosure because smaller numbers could disclose information about individuals and therefore were avoided.

3.2.3 Data

The data described in this chapter were also used in chapters 4 and 5, hence this section, Section 3.2.4 describing cost assignment, and Section 3.2.5 that describes the variables used, can be read as overarching methods for those chapters. Additional data and variables used in those chapters are described there. As the data had to be extracted from sensitive public-sector datasets, a considerable proportion of the PhD

project was spent on the application process, which involved training, ethical approval and administrative tasks. As the application process took considerably longer than expected, some datasets could not be analysed in this project but may be of use to other researchers. The datasets used in this analysis are described below.

SMR06: Cancer Registry

My cohort was derived from the Scottish Cancer Registry as recorded in SMR06. In Scotland, approximately 45,000 cancer registrations are made each year, with more than 1,400,000 recorded since 1958. The Registry records new cases in Scotland of primary malignant neoplasms, carcinoma in situ, neoplasms of uncertain behaviour, and benign brain and spinal cord tumours [3] [159]. SMR06 does not record episodes, but rather accrued information relating to a primary tumour. This information can include patient data, diagnostic information and treatment, obtained from multiple sources including computerised health records, paper records and external databases such as deaths from the General Register Office [159]. An online computer system called SOCRATES (Scottish Open Cancer Registration And Tumour Enumeration System) was developed to process the source records, which number approximately 800,000 per year. The SOCRATES database is known as SMR06. Computer validation, routine indicators, and ad hoc studies are used to assess the accuracy and completeness of data [159]. Cancer registrations are coded using the International Statistical Classification of Diseases and Related Health Problems (ICD) and the International Classification of Diseases for Oncology (ICDO) [159].

SMR00: Outpatient

An SMR00 record is generated for outpatients when they attend a medical consultant outpatient clinic, meet with a consultant outside a clinic or attend a clinic run by another healthcare professional. It can include follow-up and new attendances in all specialties excluding Accident and Emergency (A&E) and Genito-Urinary Medicine [160].

SMR01: Inpatient and Day Case

SMR01 records completed episodes of inpatient episodes and day cases in Scotland's NHS, excluding obstetric and psychiatric specialties. Episodes are specific to a specialty, and a single hospital stay for a patient may consist of multiple episodes

[3] From the year 2000 onward.

spanning multiple specialties. Over one million records are recorded annually. The ICD10 system is used to classify conditions [160].

Prescribing Information System (PIS)

The PIS dataset records medicines that are prescribed and dispensed in Scotland, with around 100 million data items per annum supplied by the Practitioner and Counter Fraud Services Division (P&CFS). It includes data on costs and drug information. The bulk of prescriptions are written by general practitioners (GPs) with the remainder written by other healthcare professionals such as nurses and dentists. It does not record prescriptions dispensed in hospitals, however, prescriptions written by hospitals and dispensed in the community are recorded [160].

Data Access, Linkage and Cleaning

Due to the highly sensitive nature of the data, all dataset linkage was performed by eDRIS in a restricted and secure environment. Records were indexed and linked by the CHI identifier, a unique number identifying all patients in NHS Scotland, then identifying information (including the CHI number) was removed and a master index file was created to index anonymised records. All subsequent analyses were performed in a secure safe haven environment with outputs thoroughly checked for potentially disclosive information by eDRIS staff before release. Publication of results with low numbers could disclose information about participants, which led me to recode some variables as described. Prior to analysis I removed duplicate records and records with a discordant death certificate, where the date of death preceded the diagnosis.

For a small proportion (<0.1%) of SMR06 records the ICD10 code could not be determined, due to duplicate records being recorded on the same day with discordant ICD10 codes and no way of knowing which code was correct, and these records were also removed from the analysis. Computational demands were an issue during the analyses and had to be minimised. The high number of records for resource use, running into tens of millions, meant that a considerable amount of cleaning and recoding was necessary. To minimise computational demands, all categories using string values were recoded to integer values. Categorical variables were aggregated unless the loss of information would have outweighed gains in computational demands. The aggregation of categories was also carried out to remove low counts that could have disclosed information about individuals.

Fourteen regional health boards were aggregated to 3 region networks. While this

entailed some loss of information, the counts in some regions were low and could have been disclosive. SIMD deciles were aggregated to quintiles. Method of first detection was aggregated to three categories: screening examination, clinical presentation, and incidental finding and other (which included not known). Stage and grade variables had complex site-specific codings across multiple variables. Where possible, stage was recoded to four categories: Stage II (early), stage III (mid), stage IV (late), not known. Due to my lack of clinical knowledge, grade variables were left untransformed and included only in models for specific cancer sites. A comorbidity was recorded if a previous SMR record with that particular comorbidity existed. If no record existed the comorbidity was recorded as absent. Where missing data were prevalent in a variable, another variable was chosen or the missing data assigned to the not known category.

Data Quality

SMR data are regularly audited in accordance with national rules and standards at Scottish Hospitals by the Information Services Division (ISD) Data Quality Assurance team. Previous audits have found that the specialty code was more than 99% accurate for 2010/11 and 2014/15, while the main condition (3-digit) was 88.3% accurate in 2010/11 and 89.0% in 2014/15 [161]. While the existence of NHS Scotland's CHI number makes novel linkages possible, it has been noted that linkage of data across databases could be improved in Scotland's health systems [16]. The reliability of demographic, diagnostic and fact of treatment has been reported as higher than that of grade, stage and treatment dates [162]. Discrepancy rates of ICD-9 codes for site were reported as 5.4% in a 1994 study while discrepancy rates for post code of residence were 7.1% [163]. For lung cancer, discrepancy rates for ICD-9 site code were found to be 4.2% in 1990 [164], while colorectal cancer had discrepancy rates of 5.5% in the same year [165]. However, the completeness of fact of treatment was not assessed and I did not find more recent data.

Patient-Public Involvement in Research

The research proposal and lay summary were presented to the Critical Care Patient-Public Involvement in Research (PPI) group in Edinburgh for feedback. The PPI group was broadly supportive of the project aims, as the topic resonated with some of their experiences during their recovery journey. They felt that the concerns related to the use of individual patient data were outweighed by the potential benefits of the research.

3.2.4 Cost Assignment

Unit costs for inpatient episodes and outpatient visits were assigned using reference costs from the Scottish Costs Book [107]. The assignment of reference costs to hospital episode statistics is believed to be a robust costing method [109]. Episodes were assigned 2017/18 unit costs using the R040 sheet for all study years. Unit costs defined in the Scottish Costs Book for inpatient episodes include resources: medical and dental, nursing, pharmacy, theatre, laboratory, and allied health practitioners (AHPs). Imputation of costs was attempted using regression modelling on PLICS costs, which were included in both SMR00 and SMR01 but with poor coverage before 2015, making them unsuitable as a primary costing method for long-term costs. Substantial errors were a problem with all regression models, causing me to reject imputation as a costing method. However, I was able to use the PLICS distributions to calibrate costs by choosing a cost-assignment method that matched desirable properties of the PLICS distributions. Costs estimated by per-episode costings had similar mean episode costs (<10% difference) to PLICS costs, whereas mean per-diem costs were approximately 30% higher than PLICS, with the standard deviation approximately doubled. However, per-diem costs better approximated median costs which may be because PLICS had many low-cost and zero-cost records. while per-episode costs contained lower variation, due to the standardised rates for recurring specialties. Despite these drawbacks, I chose per-episode costing as it better approximated the mean, which is of more interest than the median in healthcare costings, and because per-episode costings are less prone to overestimation of costs than per-diem costings [108], which was observed in this data. An alternative costing method was the use of HRGs however reference groups were only available for English NHS costs and these may not be representative of Scottish costs [108].

3.2.5 Variables

Outcome Variables

The primary outcome was total costs for an individual, equal to the sum of costs over a defined period of interest for inpatient episodes, outpatient visits and prescriptions. The defined periods of interest were the eight-year post-diagnosis period, individual years of the pre-diagnosis and post-diagnosis follow-ups, and distinct phases of care for cancer survivorship: pre-diagnosis, initial (treatment), continuing, end-of-life. Calculation of total costs required preliminary steps of measuring units of resource use and then assigning costs to units of resource use. The units of resource use were

inpatient episodes, outpatient visits and prescribed items as described below.

Inpatient episodes: Inpatient episodes were measured in the SMR01 dataset. An episode was taken to be a single spell in a facility that may span one or more days. Daycase episodes occur on a single day and were counted as one inpatient episode. A single hospital stay can consist of multiple episodes, which may require the use of two beds on a single day for a patient transferred between facilities. Continuous stays were used to calculate per-diem costs, but these were not used in the final results, hence the process will not be described. The temporal position of the episode was defined by the date of admission rather than the date of discharge, in accordance with NHS costing methods described in the Scottish Costs Book [107]. Hence if a patient was admitted on 29 December 2009 and discharged on 1 January 2010 the episode would be counted as occurring in 2009.

Outpatient visits: Outpatient visits are short units of resource use that occur on a single day in an outpatient clinic and are recorded in SMR00. Each SMR00 record describes a single outpatient visit.

Prescribed items: NHS prescriptions in Scotland are recorded in the PIS database. As the PIS dataset contained costs, I did not need to assign costs using items; however the numbers of items were of potential interest in themselves and were summed over the defined periods.

Inpatient costs: Each inpatient episode contained a specialty code, which was used to assign a cost based on reference costs for that specialty, as listed in the Scottish Costs Book [107]. Approximately 5.8% of records had no matching specialty in the Scottish Costs Book from which to apply a reference cost. These records were each given a unit cost equal to the average of the recorded specialties, which was £1262 for day cases and £4879 for episodes spanning multiple days.

Outpatient costs: The cost of a single outpatient visit was assigned a single rate of £157. Accident and emergency visits were costed at the rate of £133 for each visit. These costs were taken from the Scottish Costs Book [107].

PIS costs: Unit-cost assignment was unnecessary for prescription costs as these were provided as nominal costs in the year the prescription was made. Costs were adjusted for price inflation to 2018 price levels.

Total costs: The total costs for an individual were the sum of SMR01, SMR00 and PIS costs over a defined period, which could be yearly, one of the defined

phase-of-care phases or cumulative for the eight-year post-diagnosis follow-up. Monthly phase-of-care total costs were the total costs for a phase divided by the number of months spent in the phase. All total costs were reported in pounds sterling at 2018 price levels.

Total SMR costs: As prescription costs were not available in the pre-diagnosis period, I created an additional cost variable to describe costs without PIS costs. Trajectories including the pre-diagnosis period reported total SMR costs, whereas those covering only the post-diagnosis period reported total costs.

Explanatory Variables

Age: From the patient's age in years at the time of diagnosis, I created a categorical variable with five levels. A categorical variable was preferred to a continuous one in regression models because scatter plots suggested a non-linear relationship between age and costs. Categories were chosen to balance patient numbers while remaining intuitive and simple to interpret. The levels were: <50, 50-59, 60-69, 70-79, >=80. Additionally, a binary variable was created to indicate patients who were under 65 years of age at the time of diagnosis.

Sex: Following the coding in the SMR datasets, a person's sex was represented by a binary variable with male patients given the value 0 and female patients the value 1.

Scottish Index of Multiple Deprivation: The Scottish Index of Multiple Deprivation (SIMD) is an area-based measure of relative deprivation for people living in Scotland. To facilitate reporting and analysis [166], SIMD deciles were aggregated into quintiles represented by a categorical variable with the value 1 being the least deprived quintile and 5 the most deprived. I used SIMD 2009 version 2 because it was the most recent one available and had the highest completion rate.

Method of first detection: Listed in SMR06 was a variable describing the stage of the care pathway during which the tumour was detected. This was aggregated to 3 categories: clinical presentation, screening examination, and incidental finding / other.

Pre-diagnosis costs: These were calculated by the same method as total SMR costs, for a period of five years (5 x 365.25 rounded to 1826 days) before diagnosis to make a single value for each patient representing all SMR costs in the pre-diagnosis period.

Rurality: A variable with eight levels of urban/rural was included in my SMR01 dataset. To make it more easily interpretable and minimise low counts that could have

been disclosive, I aggregated the eight values to a binary variable representing whether an individual lived in a rural area, which was defined as a non-urban settlement with a population of less than 10,000 people or another non-urban area. Where the rurality variable was missing, the NHS region was used to impute rurality, with regions Borders, Dumfries & Galloway, Highland, Orkney, Shetland, Western Isles classed as rural on account of low population density and distance from large urban settlements.

Region network: NHS health board was coded as a text value in SMR records, representing the region code. I recoded this as a categorical variable with 14 integer categories for more efficient processing. The numbers in some regions were low, hence to avoid disclosure issues around reporting of low numbers I aggregated these categories to the three Scottish Cancer Networks: North, West, and South and East.

Stage: SMR06 contained multiple variables describing grade and stage, some site-specific and others using indexing to relate sets of values to specific sites. Completion varied between cancers, with many missing and undetermined values for some cancer sites. Variables for clinical TNM (tumour, node, metastasis) and pathological TNM stage were included but were only available for trachea, bronchus and lung, and breast cancer during 2009/10. Information on stage can be derived from the 5th digit of the ICDO3 code, or that of the ICDO2 code where no ICDO3 code was available, but these variables also had low completion. Hence information on stage could only be reliably gained for some cancers. Where stage was present, I aggregated variables to three categories to simplify reporting and increase comparability: stage II and below (localised), stage III (regional), stage IV (distant spread).

Site-specific tumour characteristics: Grade was specific to cancer sites, and therefore was only included in cancer-specific models. Additional site-specific variables were included in site-specific regression models: herceptin receptor status (HER2), side, oestrogen receptor status (erstatus). All grade and stage variable were coded as categorical variables.

Comorbidities: My SMR06 dataset lacked information on comorbidities so these had to be measured by linking pre-diagnosis SMR01 records. Fourteen comorbid conditions corresponding to components of the Charlson Comorbidity Index were represented as binary variables with 0=condition absent, 1=condition present. Where no SMR01 record existed for a patient in the period before diagnosis, all comorbidity variables for that patient were given a value of zero, based on the assumption that the patient had no condition if no hospital record for that condition existed. The 14 conditions were:

- acute myocardial infarction
- congestive heart failure
- peripheral vascular disease
- cerebral vascular disease
- dementia
- chronic pulmonary disease
- rheumatoid disease - connective tissue
- peptic ulcer
- mild liver disease
- diabetes
- diabetes with complications
- hemiplegia
- moderate or severe renal disease
- moderate or severe liver disease.

In addition to these 14 binary variables, I created a count of comorbid conditions, which was simply the sum of the 14 binary values with range [0,14].

Variables Related to Time

Variables that could vary by time were region, SIMD, comorbidities. As the study goals were to identify predictors at baseline, and also because the variables may have changed as a result of the cancer, no adjustment was made for variance of these variables over time. To assign costs to the correct period on the patient's cancer trajectory, several variables describing time were used. The date of diagnosis and date of death variables were recoded from strings to Stata's date format, as were the SMR01 admission date and the SMR00 date. Three outcome datasets were created for cumulative costs, phase-of-care costs and yearly costs. The cumulative costs dataset contained no time information. The yearly costs dataset contained a record for every patient in each year of the pre-diagnosis and post-diagnosis periods, hence was a panel dataset. If a patient used no resources in a particular year, for any reason including death, they were assigned costs of zero for that year. The phase-of-care dataset contained only records relating to resources actually used, with the particular phase

assigned to the record, as it only made sense to apply costs to phases that patients actually entered; for instance, end-of-life costs only apply to patients who died. Phases were assigned using the time since diagnosis and the time to death (if the patient died). I chose phase-of-care periods corresponding to those of Yabroff et al. (2008) [21]. Consequently, the initial, or treatment, period covered the 12 months after the diagnosis, but could be replaced by the end-of-life period if the patient died within 24 months of diagnosis. Although the initial phase can vary in length for different cancers, I followed de Oliveira (2016) [113] in using the same length for all cancers, in order to improve comparability. The end-of-life period covered the 12 months immediately preceding a patient's death, or the time from diagnosis to death if this was shorter. If a patient died during the 12 months immediately after diagnosis, all post-diagnosis costs would accrue to the end-of-life period. The continuing period applied to any time between the initial and end-of-life periods. Additionally, I included a pre-diagnosis phase, which included all time before the diagnosis up to a maximum of five years.

3.2.6 Statistical Methods

I presented sample characteristics of the cohort in tables stratified by the cancer groupings described in Section 3.2.2. Trajectories of resource use for each cancer type, and also stratified by year relative to the diagnosis, were charted. Trajectories of cumulative total costs were reported for the eight-year post-diagnosis period. Additionally, I reported trajectories of total costs and monthly total costs by phase-of-care. Total costs were the sum of inpatient, outpatient and prescriptions costs as described in Section 3.2.4. As the objective was to report resource use and associated costs for patients with cancer, rather than those specifically associated with cancer, all resource use and cost estimates were unadjusted. The reporting of phase-of-care costs followed recommendations of Wijeysundera et al. (2012) [167]. To help understand how costs related to survival, eight-year Kaplan-Meier (KM) survival plots were produced. I plotted the KM survival function rather than the hazard as this is believed to be easier to interpret [168].

In addition to charts of trajectories I produced tables of eight-year costs to show measures of precision that would be difficult to discern in stratified charts. I also produced tables of population level costs by cancer, which were simple aggregates of all patient-level costs for a particular cancer type. Prior to analysis of risk factors, unadjusted costs for subgroups of interest were charted.

To estimate the associations of risk factors on survival I used Cox proportional hazard

models on survival and GLMs on costs. GLMs comprise a linear component, a link function that is monotonic and differentiable, and dependent variables [169]. They extend the general linear model

$$y = \beta + \beta X \tag{3.1}$$

with a link function $g()$ giving the form

$$g(E(y)) = \beta + \beta X \tag{3.2}$$

where $E(y)$ is the expectation of the cost outcome, X is a vector of variables associated with y and β is a vector of parameters for X.

Advantages of GLM are that they accommodate skewness [134] and are believed to minimise prediction error [134]. While log-transformed OLS can solve some of the distributional problems with healthcare data, issues around transformation of the dependent variable are introduced [169]. GLMs avoid problems of retransformation of outcome variables because transformation is applied to the regressor rather than to the outcome itself [170]. This can also ease the interpretation of parameter coefficients [169]. A disadvantage of GLMs is that outliers can strongly affect estimates [169]. This limitation was less relevant to my analysis as the purpose of the model was not to predict outcomes, and outliers were minimised by the per-diem costing method as described in Section 3.2.4. As with OLS, the assumption of independence of observations underpins GLMs. However, the assumption of homoscedastic errors can be violated and errors need not be normally distributed [134]. The relationship between the outcome and explanatory variables need not be linear, though that between the transformed function and outcome should be [134].

To analyse risk factors for mortality I used Cox regression, which predicts the probability of death at time(t) given a baseline hazard function and patient-specific factors [168]. The measured probability was of less interest than the coefficients of the parameters, which provide estimates of the relative contribution of each predictor to the hazard of death, and were presented as exponentiated coefficients or hazard ratios (HRs). Cox regression relies on the proportional-hazards assumption, which I assessed by analysis of Schoenfeld residuals using the proportional hazards test in Stata [171].

Additionally, I visually inspected unadjusted log-log plots for categorical variables. When the curves for each group of a categorical variable are parallel, the proportional-hazards assumption is not violated [171].

Univariable analyses were performed only on all cancers combined rather than particular cancers, and the coefficients compared with those of multivariable models. Multivariable analyses were also carried out for the four most common cancer types: lung, breast, colorectal, and prostate, with coefficients for Cox and GLMs reported side-by-side for comparison. Variable selection was based on the cancer literature, expert opinion and availability of variables in the datasets. Rather than exclude insignificant variables from univariable models, I used the same variable list in multivariable models. This was in part due to the large sample numbers, which meant predictors tended to have high significance, and also to aid comparison of the coefficients and their significance for variables between GLM models and Cox models.

Although health costs tend to be dominated by zero values [133], only 136 records (<0.3%) in this analysis had zero costs, which may have resulted from the longer time frame of this analysis and also the inclusion of prescription costs—as prescriptions tend to occur more frequently than hospital visits. Model selection based on sums of squared residuals, Akaike information criterion (AIC) and Bayesian information criterion (BIC) suggested gamma regression with log-link for the distribution type and link function. As this model is an established one in health economic costing [133] I saw no reason not to use it. Raw coefficients of GLM models can be unintuitive to interpret, therefore I presented coefficients as exponents using Stata's eform option. This provided an interpretation of a coefficient as a cost ratio (CR), aiding comparison with the hazard ratio in Cox regression. The exponentiated coefficients can be interpreted as simple ratios of baseline costs, with a unit increase in the variable having a cost multiplier of β. Hence $\beta > 1$ increases costs, $\beta < 1$ decreases costs, $\beta = 1$ leaves costs unchanged. GLM coefficients were estimated by maximum likelihood. Robust standard errors were used in all GLM regressions.

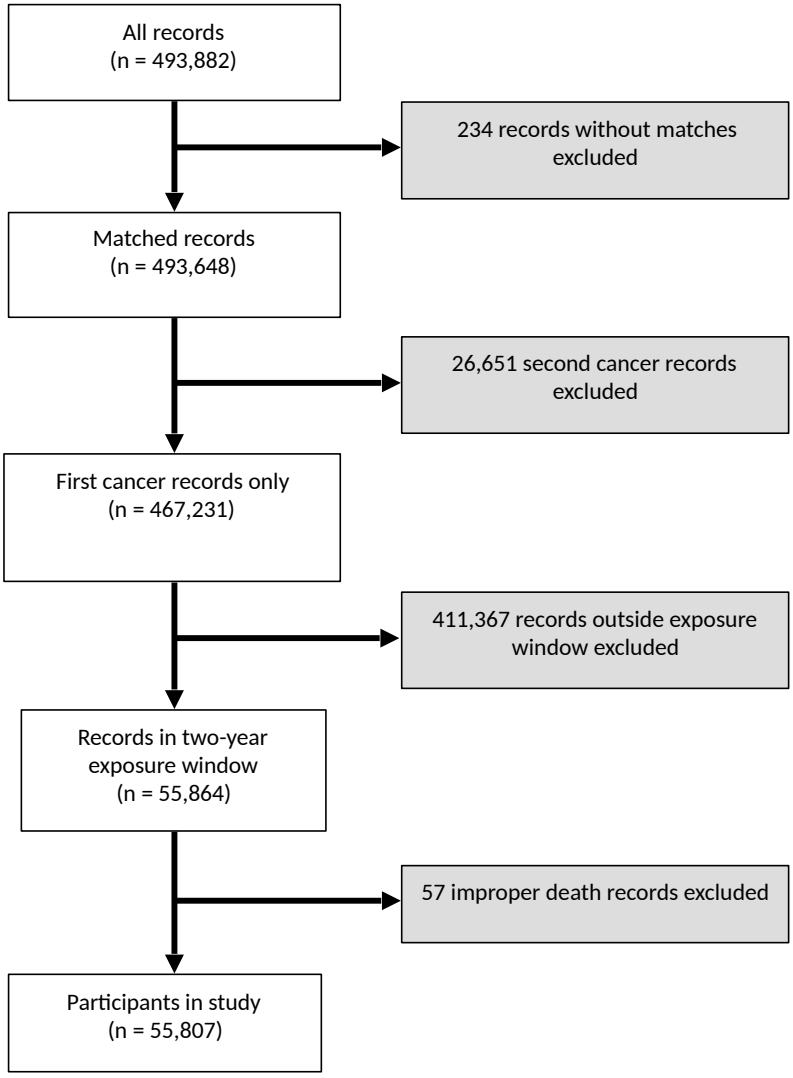
As costs described past expenditure rather than the present value of projected future expenditure, I reported all costs without discounting. Significance levels of 5% were used and 95% confidence intervals were reported where appropriate. The large sample sizes meant that the confidence intervals for costs were generally narrow and would not be visible in charts of cost trajectories. Hence they were reported only in tables of results. All analyses were carried out in Stata 15.1.

3.3 Results

3.3.1 Participants

Figure 3.1 describes the flow of study numbers from all SMR06 records in the dataset to the study participants. As described in Section 4.3.1, the cohort was created alongside a matched cancer-free cohort, which was excluded from this analysis but used in the analysis described in Chapter 4. As the matched controls were not used in this analysis they are not shown, however the matching process affected the study numbers because unmatched individuals were excluded to avoid disclosure.

Figure 3.1.: Flowchart of participant numbers



Note: Numbers prior to final selection may represent database records rather than unique individuals.

3.3.2 Descriptive Data

Tables 3.2 and 3.3 show the characteristics of the study participants. The mean age was 67.5 years old (SD 13.8 years), with the majority (62.2%) of patients being 65 years old or over. Almost half (47.2%) of patients were from the West Regional Network, with North, and South and East having 25.2% and 27.6% of patients respectively. Slightly more patients were in the more deprived quintiles than the least deprived ones: 42.5% in the second most and most deprived combined compared to 37.1% in the least and second least combined. There was considerable variation in patient characteristics across cancer sites. Within the four most common cancers, patients with trachea, bronchus and lung cancers tended to be older, come from more deprived areas, and have more comorbidities than patients with colorectal, prostate cancer and particularly breast cancer. They also tended to have a higher proportion of tumours diagnosed at an advanced stage than patients with breast cancer or colorectal cancer. Patients with head and neck cancer also tended to come from more deprived areas while skin cancer patients tended to come from less deprived areas.

Table 3.2.: Characteristics of the participants at baseline by cancer type part 1

	Trachea, bronchus and lung N=9132	Breast N=8138	Colorectal N=7270	Prostate N=5770	Head and neck N=2130	Malignant melanoma of skin N=2062
Sex						
Male	4718 (51.7%)	46 (0.6%)	3942 (54.2%)	5770 (100.0%)	1481 (69.5%)	905 (43.9%)
Female	4414 (48.3%)	8092 (99.4%)	3328 (45.8%)	NA	649 (30.5%)	1157 (56.1%)
Age in years (mean, sd)	71.4 (10.6)	62.6 (14.1)	70.4 (11.8)	70.6 (9.4)	63.7 (12.3)	59.5 (17.5)
Age						
< 50	249 (2.7%)	1551 (19.1%)	325 (4.5%)	55 (1.0%)	249 (11.7%)	635 (30.8%)
50-59	996 (10.9%)	1896 (23.3%)	952 (13.1%)	599 (10.4%)	500 (23.5%)	355 (17.2%)
60-69	2396 (26.2%)	2191 (26.9%)	1909 (26.3%)	2030 (35.2%)	683 (32.1%)	396 (19.2%)
70-79	3334 (36.5%)	1396 (17.2%)	2379 (32.7%)	2026 (35.1%)	479 (22.5%)	400 (19.4%)
>= 80	2157 (23.6%)	1100 (13.5%)	1701 (23.4%)	1059 (18.4%)	219 (10.3%)	276 (13.4%)
Under 65 years old	2247 (24.6%)	4619 (56.8%)	2162 (29.7%)	1588 (27.5%)	1121 (52.6%)	1202 (58.3%)
SIMD quintile						
Least deprived	1005 (11.0%)	1757 (21.6%)	1389 (19.1%)	1306 (22.6%)	247 (11.6%)	559 (27.1%)
2nd least deprived	1339 (14.7%)	1755 (21.6%)	1437 (19.8%)	1273 (22.1%)	350 (16.4%)	448 (21.7%)
3rd least deprived	1734 (19.0%)	1660 (20.4%)	1522 (20.9%)	1238 (21.5%)	409 (19.2%)	435 (21.1%)
2nd most deprived	2256 (24.7%)	1568 (19.3%)	1553 (21.4%)	1089 (18.9%)	488 (22.9%)	338 (16.4%)
Most deprived	2798 (30.6%)	1398 (17.2%)	1369 (18.8%)	864 (15.0%)	636 (29.9%)	282 (13.7%)
Regional network						
South and east	2463 (27.0%)	2168 (26.6%)	2032 (28.0%)	1752 (30.4%)	564 (26.5%)	540 (26.2%)
West	4684 (51.3%)	3820 (46.9%)	3284 (45.2%)	2498 (43.3%)	1089 (51.1%)	997 (48.4%)
North	1985 (21.7%)	2150 (26.4%)	1954 (26.9%)	1520 (26.3%)	477 (22.4%)	525 (25.5%)
Patient recorded as dying *	8638 (94.6%)	2622 (32.2%)	4313 (59.3%)	2570 (44.5%)	1266 (59.4%)	552 (26.8%)
Survival months (mean, sd)	15.9 (24.5)	77.7 (30.8)	54.3 (39.9)	71.2 (33.4)	56.1 (39.1)	81.4 (27.9)
Stage II and below	2793 (30.6%)	6746 (82.9%)	1146 (15.8%)	No info	No info	No info
Stage IV	4501 (49.3%)	501 (6.2%)	1276 (17.6%)	No info	No info	No info
Method of 1st detection						
Screening examination	NA	2555 (31.4%)	1002 (13.8%)	NA	NA	NA
Clinical presentation	8606 (94.2%)	5282 (64.9%)	6120 (84.2%)	5513 (95.5%)	2106 (98.9%)	2042 (99.0%)
Incidental finding and other	526 (5.8%)	301 (3.7%)	148 (2.0%)	257 (4.5%)	24 (1.1%)	20 (1.0%)
Comorbidity count						
Zero	6440 (70.5%)	7415 (91.1%)	6167 (84.8%)	5058 (87.7%)	1796 (84.3%)	1901 (92.2%)
One	1996 (21.9%)	577 (7.1%)	853 (11.7%)	547 (9.5%)	253 (11.9%)	125 (6.1%)
Two or more	696 (7.6%)	146 (1.8%)	250 (3.4%)	165 (2.9%)	81 (3.8%)	36 (1.7%)
Comorbidities						
Acute myocardial infarcation	266 (2.9%)	80 (1.0%)	135 (1.9%)	124 (2.1%)	50 (2.3%)	25 (1.2%)
Congestive heart failure	192 (2.1%)	67 (0.8%)	104 (1.4%)	55 (1.0%)	14 (0.7%)	19 (0.9%)
Peripheral vascular disease	292 (3.2%)	49 (0.6%)	103 (1.4%)	83 (1.4%)	<30	<30
Cerebral vascular disease	297 (3.3%)	82 (1.0%)	104 (1.4%)	118 (2.0%)	56 (2.6%)	18 (0.9%)
Dementia	168 (1.8%)	83 (1.0%)	97 (1.3%)	54 (0.9%)	16 (0.8%)	13 (0.6%)
Chronic pulmonary disease	1404 (15.4%)	226 (2.8%)	279 (3.8%)	172 (3.0%)	116 (5.4%)	48 (2.3%)
Rheumatoid disease	101 (1.1%)	35 (0.4%)	38 (0.5%)	14 (0.2%)	<10	<10
Peptic ulcer	40 (0.4%)	24 (0.3%)	33 (0.5%)	11 (0.2%)	12 (0.6%)	<10
Mild liver disease	52 (0.6%)	22 (0.3%)	43 (0.6%)	12 (0.2%)	28 (1.3%)	<10
Diabetes	431 (4.7%)	151 (1.9%)	314 (4.3%)	154 (2.7%)	69 (3.2%)	39 (1.9%)
Diabetes with complications	14 (0.2%)	<10	<10	<10	<10	<10
Hemiplegia	30 (0.3%)	<10	11 (0.2%)	11 (0.2%)	<10	<10
Renal disease - moderate or severe	209 (2.3%)	63 (0.8%)	112 (1.5%)	89 (1.5%)	12 (0.6%)	10 (0.5%)
Liver disease - moderate or severe	15 (0.2%)	<10	<10	<10	<10	<10

Notes: SIMD = Scottish Index of Multiple Deprivation, sd = standard deviation.
Comorbidity measures were recorded prior to SMR06 registration and do not include cancers.
Counts are rounded to <10 and <30 to prevent differencing across rows and columns disclosing patient information.
* Death may have occurred after the eight-year follow-up.

Table 3.3.: Characteristics of the participants at baseline by cancer type part 2

	Kidney N=1541	Non-Hodgkin lymphoma N=1881	Oesophagus N=1590	Bladder N=1373	All other cancers N=14920	All cancers N=55807
Sex						
Male	877 (56.9%)	932 (49.5%)	1004 (63.1%)	910 (66.3%)	6350 (42.6%)	26935 (48.3%)
Female	664 (43.1%)	949 (50.5%)	586 (36.9%)	463 (33.7%)	8570 (57.4%)	28872 (51.7%)
Age in years (mean, sd)	67.6 (12.9)	66.3 (14.2)	70.2 (11.6)	73.8 (10.9)	66.2 (16.1)	67.5 (13.8)
Age						
< 50	140 (9.1%)	229 (12.2%)	67 (4.2%)	35 (2.5%)	2348 (15.7%)	5883 (10.5%)
50-59	261 (16.9%)	316 (16.8%)	226 (14.2%)	103 (7.5%)	2015 (13.5%)	8219 (14.7%)
60-69	417 (27.1%)	487 (25.9%)	446 (28.1%)	303 (22.1%)	3351 (22.5%)	14609 (26.2%)
70-79	428 (27.8%)	504 (26.8%)	481 (30.3%)	463 (33.7%)	3969 (26.6%)	15859 (28.4%)
>= 80	295 (19.1%)	345 (18.3%)	370 (23.3%)	469 (34.2%)	3237 (21.7%)	11228 (20.1%)
Under 65 years	628 (40.8%)	805 (42.8%)	513 (32.3%)	263 (19.2%)	5951 (39.9%)	21099 (37.8%)
SIMD quintile						
Least deprived	274 (17.8%)	392 (20.8%)	234 (14.7%)	238 (17.3%)	2627 (17.6%)	10028 (18.0%)
2nd least deprived	274 (17.8%)	368 (19.6%)	289 (18.2%)	248 (18.1%)	2871 (19.2%)	10652 (19.1%)
3rd least deprived	331 (21.5%)	404 (21.5%)	337 (21.2%)	303 (22.1%)	3061 (20.5%)	11434 (20.5%)
2nd most deprived	343 (22.3%)	377 (20.0%)	355 (22.3%)	288 (21.0%)	3228 (21.6%)	11883 (21.3%)
Most deprived	319 (20.7%)	340 (18.1%)	375 (23.6%)	296 (21.6%)	3133 (21.0%)	11810 (21.2%)
Regional network						
South and east	433 (28.1%)	559 (29.7%)	403 (25.3%)	387 (28.2%)	4101 (27.5%)	15402 (27.6%)
West	715 (46.4%)	818 (43.5%)	761 (47.9%)	603 (43.9%)	7099 (47.6%)	26368 (47.2%)
North	393 (25.5%)	504 (26.8%)	426 (26.8%)	383 (27.9%)	3720 (24.9%)	14037 (25.2%)
Patient recorded as dying *	921 (59.8%)	955 (50.8%)	1464 (92.1%)	1024 (74.6%)	10519 (70.5%)	34844 (62.4%)
Survival months (mean, sd)	52.9 (40.7)	60.6 (40.2)	19.8 (26.7)	41.4 (38.0)	39.3 (40.7)	49.1 (41.2)
Stage II and below	No info	No info	No info	No info	296 (2.0%)	10981 (19.7%)
Stage IV	No info	No info	No info	No info	222 (1.5%)	6500 (11.6%)
Method of 1st detection						
Screening examination	NA	NA	NA	NA	NA	3557 (6.4%)
Clinical presentation	1325 (86.0%)	1814 (96.4%)	1565 (98.4%)	1340 (97.6%)	13772 (92.3%)	49485 (88.7%)
Incidental finding and other	216 (14.0%)	67 (3.6%)	25 (1.6%)	33 (2.4%)	1148 (7.7%)	2765 (5.0%)
Comorbidity count						
Zero	1243 (80.7%)	1565 (83.2%)	1295 (81.4%)	1101 (80.2%)	12000 (80.4%)	45981 (82.4%)
One	218 (14.1%)	239 (12.7%)	237 (14.9%)	201 (14.6%)	2247 (15.1%)	7493 (13.4%)
Two or more	80 (5.2%)	77 (4.1%)	58 (3.6%)	71 (5.2%)	673 (4.5%)	2333 (4.2%)
Acute myocardial infarcation	43 (2.8%)	34 (1.8%)	35 (2.2%)	37 (2.7%)	287 (1.9%)	1116 (2.0%)
Congestive heart failure	42 (2.7%)	26 (1.4%)	22 (1.4%)	22 (1.6%)	230 (1.5%)	793 (1.4%)
Peripheral vascular disease	31 (2.0%)	<30	<30	<30	186 (1.2%)	867 (1.6%)
Cerebral vascular disease	24 (1.6%)	29 (1.5%)	18 (1.1%)	22 (1.6%)	311 (2.1%)	1079 (1.9%)
Dementia	23 (1.5%)	18 (1.0%)	22 (1.4%)	23 (1.7%)	250 (1.7%)	767 (1.4%)
Chronic pulmonary disease	60 (3.9%)	74 (3.9%)	92 (5.8%)	79 (5.8%)	676 (4.5%)	3226 (5.8%)
Rheumatoid disease	10 (0.6%)	19 (1.0%)	<10	<10	103 (0.7%)	349 (0.6%)
Peptic ulcer	<10	20 (1.1%)	19 (1.2%)	<10	159 (1.1%)	334 (0.6%)
Mild liver disease	16 (1.0%)	20 (1.1%)	19 (1.2%)	<10	294 (2.0%)	517 (0.9%)
Diabetes	69 (4.5%)	75 (4.0%)	61 (3.8%)	72 (5.2%)	735 (4.9%)	2170 (3.9%)
Diabetes with complications	<10	<10	<10	<10	13 (0.1%)	65 (0.1%)
Hemiplegia	<10	<10	<10	<10	47 (0.3%)	139 (0.2%)
Renal disease - moderate or severe	56 (3.6%)	45 (2.4%)	23 (1.4%)	55 (4.0%)	317 (2.1%)	991 (1.8%)
Liver disease - moderate or severe	<10	<10	<10	<10	114 (0.8%)	166 (0.3%)

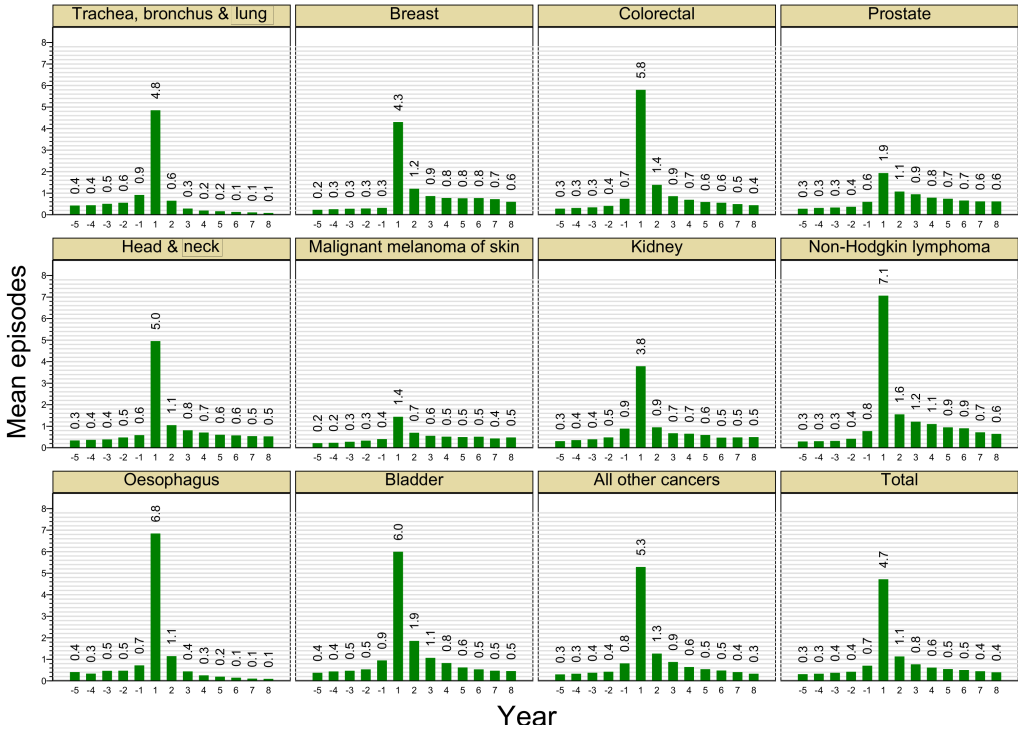
Notes: SIMD = Scottish Index of Multiple Deprivation, sd = standard deviation.
Comorbidity measures were recorded prior to SMR06 registration and do not include cancers.
Counts are rounded to <10 and <30 to prevent differencing across rows and columns disclosing patient information.
* Death may have occurred after the eight-year follow-up.

3.3.3 Trajectories of Resource Use and Costs

The trajectories in figures 3.2 to 3.4 show several common features across cancer types and resource types. There is a prominent spike in the first year after diagnosis, then a sharp drop in year 2 and more gradual declines thereafter. Inpatient episodes

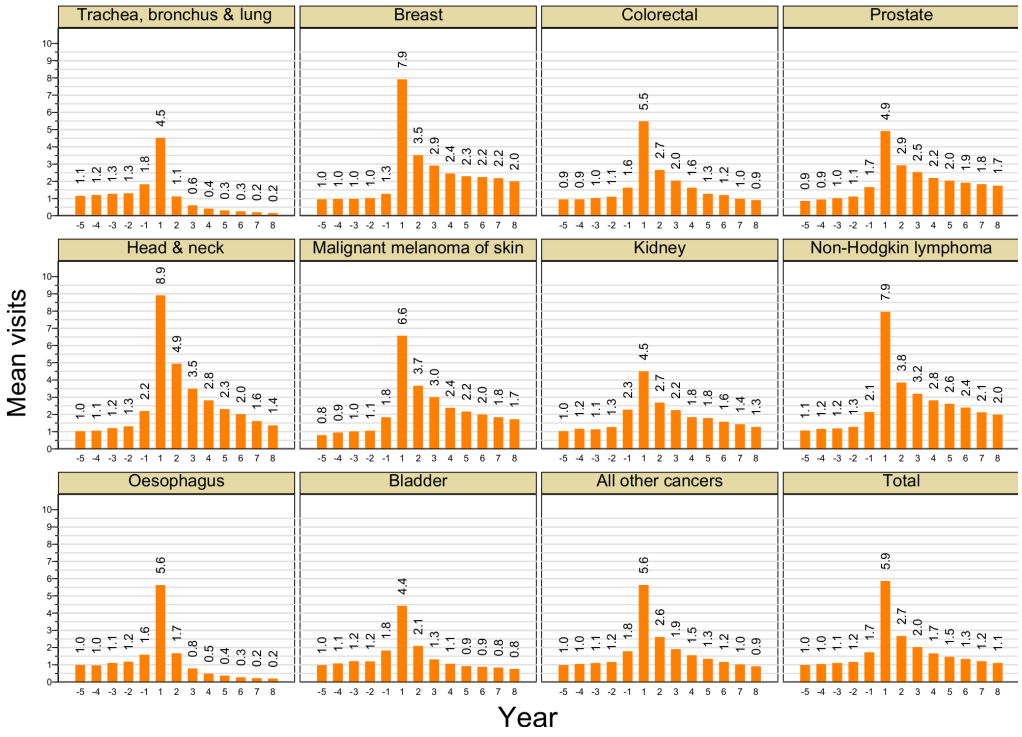
tended to be more peaked in year 1 than outpatient visits or prescribed items, though the latter was only visible in the post-diagnosis period. Care should be taken when comparing the pre-diagnosis period to the post-diagnosis period, because all participants had to be alive before but not after diagnosis. Additionally, data for the pre-diagnosis period were not available for prescribed items hence are not shown in Figure 3.4. Although hospital use after diagnosis would be reduced by mortality, it remained above pre-diagnosis levels for several cancers.

Figure 3.2.: Trajectories of inpatient episodes by year and by cancer type



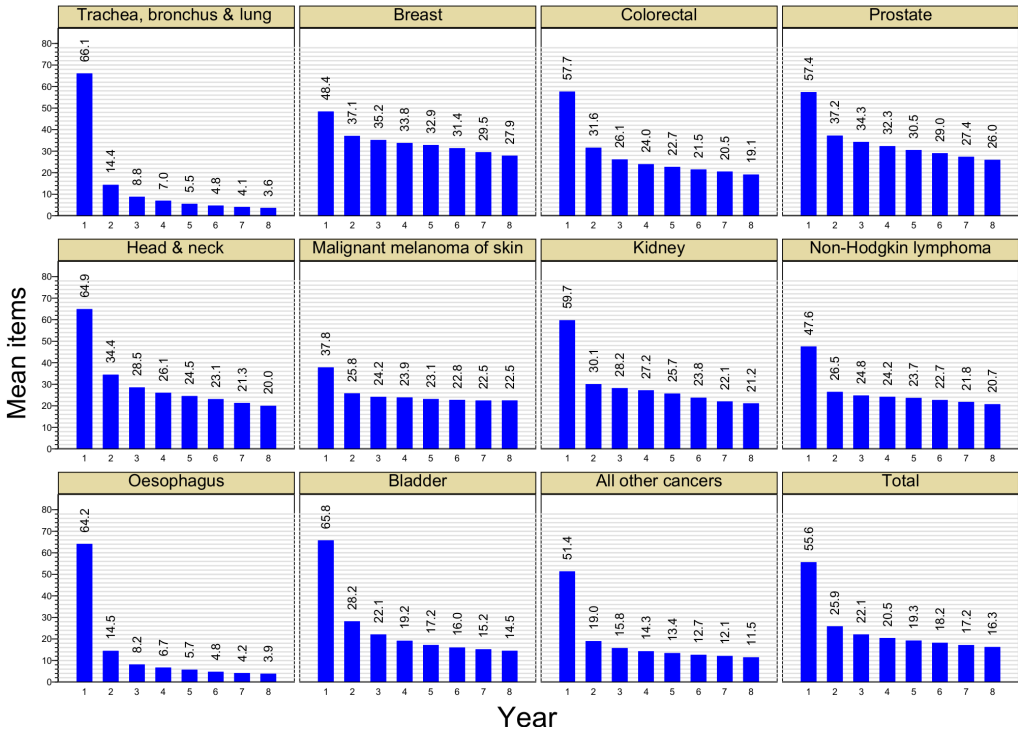
Notes: All participants were represented in all years. Years are relative to the diagnosis.

Figure 3.3.: Trajectories of outpatient visits by year and by cancer type



Notes: All participants were represented in all years. Years are relative to the diagnosis.

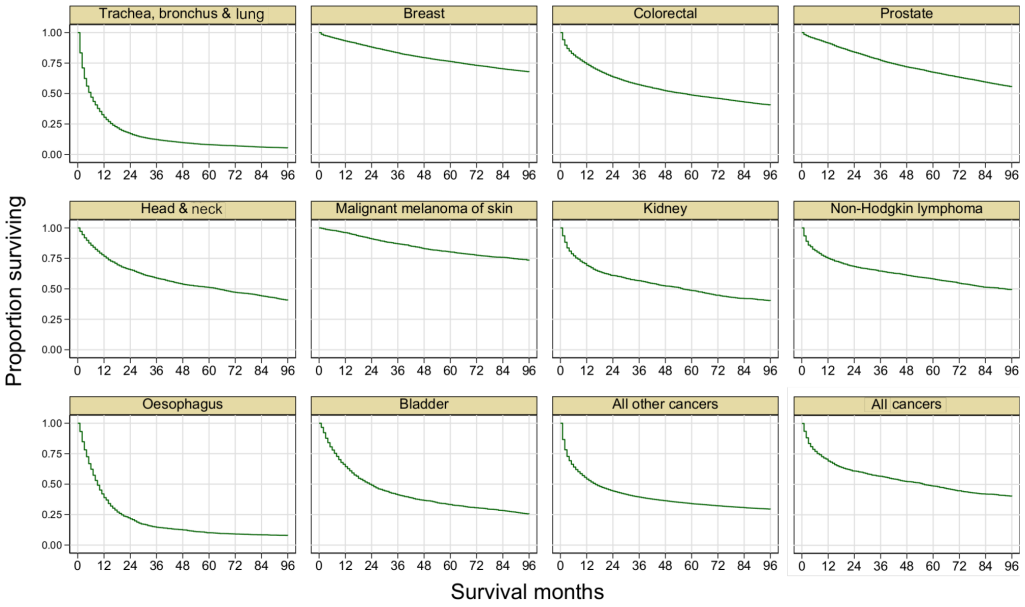
Figure 3.4.: Trajectories of prescribed items by year and by cancer type



Notes: All participants were represented in all years. Years are relative to the diagnosis.
The pre-diagnosis period is not shown as prescriptions data were lacking.

Figure 3.5 shows survival rates over time for the different cancers. Approximately two-thirds of trachea, bronchus and lung cancer patients died within twelve months of diagnosis and less than 5% survived the eight-year follow-up. Oesophagus cancer had similarly low survival whereas around three-quarters of malignant melanoma of skin patients and a similar proportion of breast cancer patients survived to the end of the eight-year period. Other cancers fell between these extremes. There was also variation in the slopes of the curves and how these changed over time; cancers with higher mortality tended to have very steep slopes in the first twelve months that became less steep over time, and towards the end of the eight-year follow up the curves for all cancers were relatively flat with similar slopes, though at different levels.

Figure 3.5.: Eight-year Kaplan-Meier post-diagnosis survival trajectories by cancer type

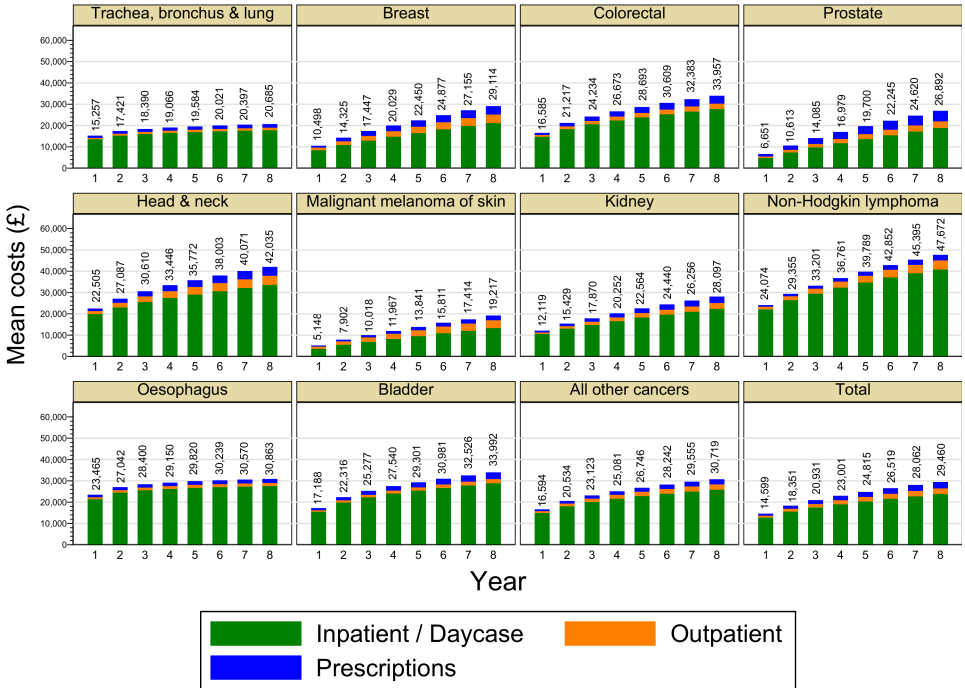


The differing survival trajectories across cancer types appear to have been reflected in the trajectories of resource use. Cancers with the highest mortality rates had high resource use in the diagnosis year, but resource use then dipped sharply and fell below pre-diagnosis levels. Cancers with high survival such as skin cancer had relatively low resource use in the diagnosis year, but the falls in subsequent years were less marked than cancers with high mortality. Cancers with moderate severity, such as kidney cancer, colorectal cancer and non-Hodgkin lymphoma had moderate to high resource use in the diagnosis year which stayed elevated above pre-diagnosis levels.

The impact of this relationship on overall costs can be observed in figures 3.6–3.8, which show cumulative costs and costs by phase-of-care. High-mortality cancers had high monthly rates of costs in the initial and end-of-life phases, which contributed to

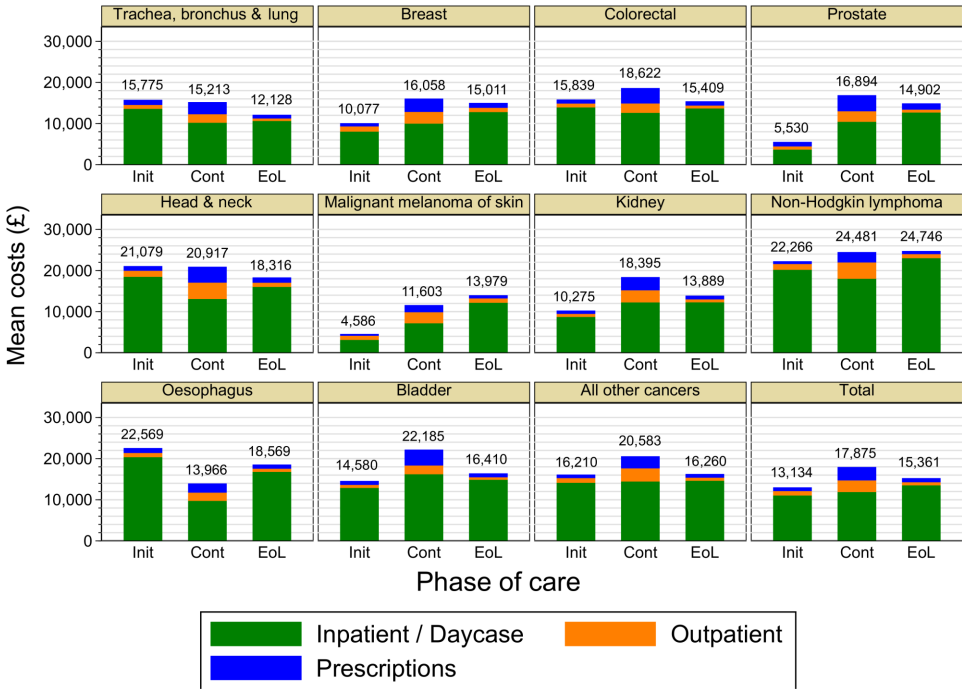
moderately high cumulative costs in year 1. Costs then accumulated more gradually due to lower survival. Cancers with high survival such as skin and prostate had low costs in the initial phase and moderate costs in other phases, leading to low costs in the diagnosis year that then rose steadily throughout the eight-year period, resulting in moderate overall costs. Cancers with moderate survival, such as head and neck cancer, had moderate to high costs in all phases. year 1 costs were moderately high and thereafter rose steadily, leading to the highest costs. While considerable variation between rates of cost accumulation can be seen in Figure 3.8, all cancers had the highest rate during the end-of-life phase and the lowest rate in the continuing phase. Cancers with higher mortality tended to have higher monthly costs.

Figure 3.6.: Trajectories of cumulative costs by year and by cancer type



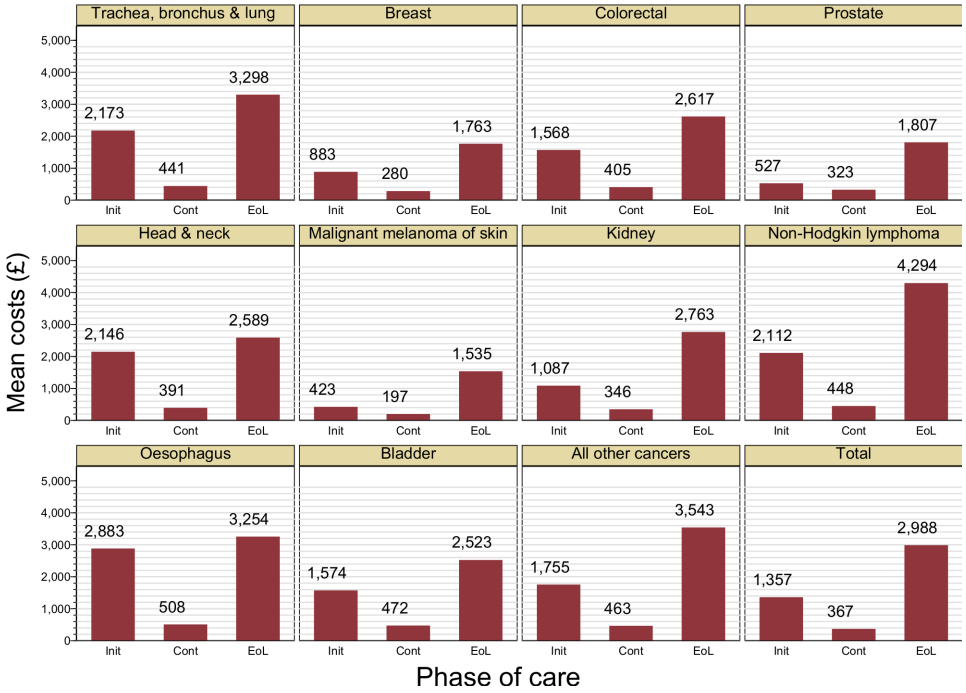
Notes: Years are relative to the diagnosis. All participants were represented in all years. All costs are undiscounted at 2018 price levels.

Figure 3.7.: Trajectories of costs by phase-of-care and by cancer type



Notes: Init = initial, Cont = continuing, EoL = end of life. Only participants who entered a phase were represented in that phase. All costs are undiscounted at 2018 price levels.

Figure 3.8.: Trajectories of total monthly costs by phase-of-care and by cancer type



Notes: Init = initial, Cont = continuing, EoL = end of life. Only participants who entered a phase were represented in that phase. Total costs were the sum of inpatient/daycase, outpatient and prescriptions. All costs are undiscounted at 2018 price levels.

3.3.4 Eight-year cumulative costs

The cost breakdowns in figures 3.6 to 3.7 show that inpatient costs were the largest component of costs in all post-diagnosis periods and phases, however their relative magnitude was smaller in the continuous phase. Outpatient and prescription costs appear to have had approximately similar magnitudes. Additional eight-year, phase-of-care and yearly costs are described in tables 3.4 to 3.5. Table 3.4 shows the eight-year costs for men and women. The highest eight-year costs were observed in non-Hodgkin lymphoma (mean £47,672; 95%CI £45,500 to £49,843), head and neck cancer (mean £42,035; 95%CI £40,555 to £43,515) and colorectal cancer (mean £33,957; 95%CI £33,265 to £34,649). The lowest costs were observed in malignant melanoma of skin (mean £19,217; 95%CI £18,251 to £20,184) trachea, bronchus and lung cancers (mean £20,685; 95%CI £20,263 to 21,107) and prostate cancer (mean £26,892; 95%CI £26,299 to £27,485). Costs for males were higher for all applicable cancers except trachea, bronchus and lung cancers (mean £20,155; 95%CI £19,578 to £20,732) compared to (mean £21,252; 95%CI £20,634 to £21,869) for women.

Table 3.4.: Eight-year cumulative mean total costs by cancer type and by sex

	Cancer type	Mean (£)	95% CI (£)	
All persons	Trachea, bronchus & lung	20,685	20,263	21,107
	Breast	29,114	28,508	29,721
	Colorectal	33,957	33,265	34,649
	Prostate	26,892	26,299	27,485
	Head & neck	42,035	40,555	43,515
	Malignant melanoma of skin	19,217	18,251	20,184
	Kidney	28,097	26,741	29,454
	Non-Hodgkin lymphoma	47,672	45,500	49,843
	Oesophagus	30,863	29,596	32,131
	Bladder	33,992	32,646	35,339
	All other cancers	30,719	30,103	31,334
Men	Trachea, bronchus & lung	20,155	19,578	20,732
	Breast	35,333	24,320	46,346
	Colorectal	35,592	34,641	36,543
	Prostate	26,892	26,299	27,485
	Head & neck	42,422	40,607	44,237
	Malignant melanoma of skin	20,431	18,965	21,896
	Kidney	27,818	26,151	29,484
	Non-Hodgkin lymphoma	51,039	47,558	54,520
	Oesophagus	33,377	31,720	35,034
	Bladder	34,829	33,159	36,500
	All other cancers	32,681	31,654	33,709
Women	Trachea, bronchus & lung	21,252	20,634	21,869
	Breast	29,079	28,472	29,686
	Colorectal	32,020	31,015	33,026
	Prostate	NA	NA	NA
	Head & neck	41,151	38,615	43,687
	Malignant melanoma of skin	18,268	16,984	19,552
	Kidney	28,466	26,215	30,718
	Non-Hodgkin lymphoma	44,364	41,766	46,963
	Oesophagus	26,557	24,665	28,448
	Bladder	32,347	30,078	34,616
	All other cancers	29,265	28,511	30,018

Notes: Costs are the sum of inpatient/daycase, outpatient and prescriptions. Years are relative to the diagnosis. All costs are undiscounted at 2018 price levels. CI = confidence interval. NA = not applicable.

The overall cost of resources used for all patients over the eight-year follow-up is shown in Table 3.5. For all cancers combined the total cost of resources used was £1,644,059,782. The highest overall costs were observed in colorectal at £246,866,814, then breast at £236,932,215 followed by trachea, bronchus and lung at £188,897,080 and prostate at £155,167,399.

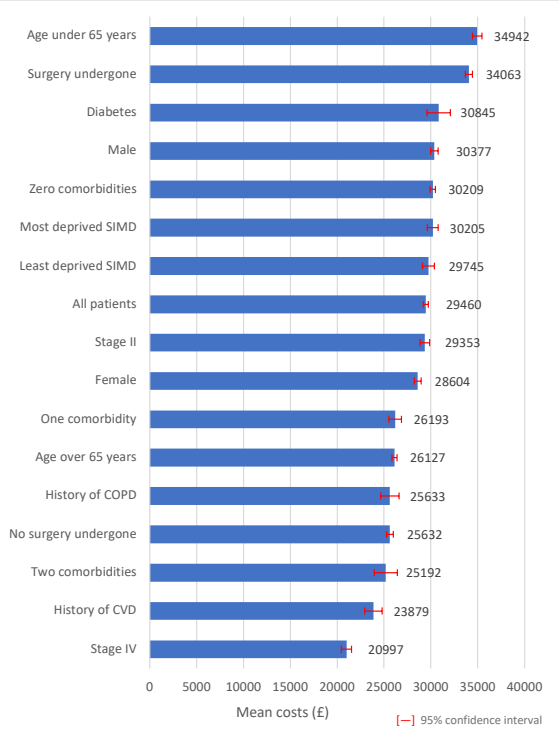
Table 3.5.: Eight-year population-level total costs by cancer type

	Sum all patients (£)			
Cancer type	**Outpatient**	**Prescriptions**	**Inpatient**	**Total costs**
Trachea, bronchus & Lung	10,811,066	15,290,002	162,796,012	188,897,080
Breast	32,409,633	32,011,976	172,510,606	236,932,215
Colorectal	18,382,498	26,699,508	201,784,808	246,866,814
Prostate	18,130,270	28,362,201	108,674,928	155,167,399
Head & neck	9,171,880	8,791,090	71,571,670	89,534,640
Malignant melanoma of skin	7,531,563	4,512,761	27,581,509	39,625,833
Kidney	4,182,416	4,676,768	34,438,676	43,297,860
Non-Hodgkin lymphoma	7,938,148	4,907,595	76,824,436	89,670,179
Oesophagus	2,396,500	2,919,618	43,756,840	49,072,958
Bladder	2,633,992	4,309,894	39,727,654	46,671,540
All other cancers	37,620,147	35,187,827	385,515,290	458,323,264
All cancers	151,208,113	167,669,240	1,325,182,429	1,644,059,782

Note: All costs are undiscounted at 2018 price levels. Costs are the aggregate of all patients combined for each cancer type.

Figure 3.9 shows eight-year total costs over all cancers for sub-groups. Higher costs were measured for males, under 65s and those without comorbidities, while patients with stage IV cancers had substantially lower costs.

Figure 3.9.: Eight-year mean total costs for subgroups



Notes: SIMD = Scottish Index of Multiple Deprivation, COPD = chronic obstructive pulmonary disease, CVD = cardiovascular disease. Costs are the sum of inpatient/daycase, outpatient and prescriptions. Costs are per-person. All costs are undiscounted at 2018 price levels.

3.3.5 Analysis of Risk Factors for Costs and Mortality

Table 3.6 compares results on all patients from GLM models on costs with Cox proportional hazard models on survival, and univariable models with multivariable ones. Notable associations were found with being aged over 80 years old (adjusted CR 0.549; $p<0.001$, adjusted HR 5.942; $p<0.001$), stage IV (adjusted CR 0.691; $p<0.001$, adjusted HR 3.316; $p<0.001$), and dementia (adjusted CR 0.513, $p<0.001$, adjusted HR 1.687; $p<0.001$) The general pattern of associations was that factors positively associated with mortality were negatively associated with costs, and vice versa. However, there were exceptions. Factors with significant positive associations between both costs and mortality were pre-diagnosis costs and diabetes, while female sex, screening, and rurality had negative associations with both costs and hazard of death, although rural was only significant at the 5% level in univariable models. Factors negatively associated with costs and positively associated with hazard of death included age, stage IV and many comorbidities, most notably dementia. Associations tended to have lower magnitudes in multivariable models. Cox models showed significant associations between SIMD quintile and mortality, with higher deprivation being positively associated with higher mortality. However, there was no clear association between SIMD quintile and costs, with only the second least deprived quintile being significant at the 5% level and only small differences in the magnitude of associations between quintiles. Further multivariable models are shown in tables 3.7 to 3.10 for the four most common cancer types.

Table 3.6.: Univariable and multivariable results for GLM regression on costs and Cox regression on hazard of death for all cancer patients over the eight-year post-diagnosis period

	Univariable				Multivariable			
	GLM		Cox		GLM		Cox	
Variable	**CR**	***p***	**HR**	***p***	**CR**	***p***	**HR**	***p***
Age								
< 50	1	Reference	1	Reference	1	Reference	1	Reference
50-59	0.973	0.157	1.820	<0.001	0.998	0.929	1.914	<0.001
60-69	0.893	<0.001	2.545	<0.001	0.908	<0.001	2.511	<0.001
70-79	0.774	<0.001	4.237	<0.001	0.778	<0.001	3.706	<0.001
>= 80	0.539	<0.001	7.359	<0.001	0.549	<0.001	5.942	<0.001
Female	0.942	<0.001	0.819	<0.001	0.962	<0.001	0.975	0.022
SIMD								
least deprived	1	Reference	1	Reference	1	Reference	1	Reference
2nd least deprived	0.963	0.013	1.159	<0.001	0.971	0.042	1.152	<0.001
3rd least deprived	0.973	0.060	1.279	<0.001	0.980	0.139	1.220	<0.001
2nd most deprived	0.999	0.923	1.470	<0.001	0.999	0.948	1.332	<0.001
most deprived	1.015	0.295	1.663	<0.001	0.998	0.889	1.505	<0.001
Method of 1st detection								
clinical presentation	1	<0.001	1	Reference	1	Reference	1	Reference
screening examination	0.929	<0.001	0.170	<0.001	0.851	<0.001	0.304	<0.001
incidental finding and other	0.886	<0.001	1.005	0.840	0.907	<0.001	0.965	0.138
Pre-diagnosis costs (per £1000)	1.003	<0.001	1.017	<0.001	1.006	<0.001	1.009	<0.001
Rural	0.983	<0.001	0.877	<0.001	0.992	0.430	0.954	<0.001
Region Network								
west	1	Reference	1	Reference	1	Reference	1	Reference
south and east	1.001	0.949	0.916	<0.001	1.021	0.057	0.958	0.002
north	0.903	<0.001	0.948	<0.001	0.932	<0.001	0.982	0.215
Stage								
II or below (early)	0.995	<0.001	0.433	<0.001	0.992	0.487	0.608	<0.001
IV (late)	0.687	<0.001	3.575	<0.001	0.691	<0.001	3.136	<0.001
Acute myocardial infarction	0.959	0.269	1.743	<0.001	1.038	0.281	1.077	0.035
Congestive heart failure	0.679	<0.001	2.588	<0.001	0.795	<0.001	1.334	<0.001
Peripheral vascular disease	0.843	<0.001	2.166	<0.001	0.911	0.003	1.272	<0.001
Cerebral vascular disease	0.699	<0.001	2.190	<0.001	0.811	<0.001	1.263	<0.001
Dementia	0.381	<0.001	3.496	<0.001	0.513	<0.001	1.687	<0.001
Chronic pulmonary disease	0.862	<0.001	2.119	<0.001	0.912	<0.001	1.416	<0.001
Rheumatoid disease - Connective tissue	1.027	0.596	1.643	<0.001	1.055	0.260	1.205	0.002
Peptic ulcer	1.027	0.619	1.640	<0.001	1.036	0.480	1.294	<0.001
Mild liver disease	0.978	0.636	1.905	<0.001	0.892	0.013	1.770	<0.001
Diabetes	1.044	0.043	1.761	<0.001	1.103	<0.001	1.206	<0.001
Diabetes with complications	1.240	0.100	1.717	<0.001	1.220	0.136	1.208	0.214
Hemiplegia	0.960	0.651	2.105	<0.001	0.998	0.983	1.497	<0.001
Moderate or severe renal disease	0.945	0.245	2.401	<0.001	1.037	0.366	1.348	<0.001
Moderate or severe liver disease	0.926	0.389	2.116	<0.001	0.828	0.041	1.508	<0.001

Abbreviations: GLM = generalised linear model, CR = cost ratio, HR = hazard ratio, SIMD = Scottish Index of Multiple Deprivation

Table 3.7 shows the results of multivariable regression for trachea, bronchus and lung cancer. Associations for age and stage were similar to those in multivariable models with all patients, however, female sex was positively associated with higher costs (HR 1.086; p<0.001) while higher deprivation was significantly associated with lower costs and with higher mortality. However, the magnitudes of the associations were lower than those with age and stage. Grade was also associated with both costs and mortality but no clear pattern was apparent.

Table 3.7.: Results of multivariable Cox regression on mortality risk and multivariable GLM regression on total eight-year post-diagnosis costs for trachea, bronchus and lung patients

	GLM		Cox	
Variable	**CR**	***p***	**HR**	***p***
Age				
< 50	1	Reference	1	Reference
50-59	0.958	0.401	1.390	<0.001
60-69	0.809	<0.001	1.514	<0.001
70-79	0.614	<0.001	1.885	<0.001
>= 80	0.437	<0.001	2.477	<0.001
Female	1.086	<0.001	0.886	<0.001
SIMD				
least deprived	1	Reference	1	Reference
2nd least deprived	0.960	0.266	1.106	0.016
3rd least deprived	0.931	0.034	1.080	0.050
2nd most deprived	0.921	0.010	1.111	0.004
most deprived	0.904	0.002	1.199	<0.001
Method of 1st detection				
clinical presentation	1	Reference	1	Reference
incidental finding and other	1.148	0.002	0.796	<0.001
Pre-diagnosis costs (per £1000)	1.005	<0.001	1.003	<0.001
Rural	0.944	0.010	1.012	0.638
Region network				
west	1	Reference	1	Reference
south and east	1.065	0.009	1.074	0.009
north	1.032	0.215	1.015	0.626
Stage				
localised	1	Reference	1	Reference
regional spread	0.754	<0.001	1.891	<0.001
distant metastasis	0.505	<0.001	3.818	<0.001
not known	0.680	<0.001	1.999	<0.001
Grade				
1	1	Reference	1	Reference
2	0.963	0.647	1.152	0.131
3	0.817	0.015	1.655	<0.001
4	0.817	0.064	2.002	<0.001
not known	0.648	<0.001	2.070	<0.001
Acute myocardial infarction	1.129	0.035	0.922	0.171
Congestive heart failure	0.743	<0.001	1.364	<0.001
Peripheral vascular disease	0.869	0.005	1.111	0.056
Cerebral vascular disease	0.842	0.002	1.119	0.057
Dementia	0.616	<0.001	1.158	0.073
Chronic pulmonary disease	0.934	0.008	1.121	<0.001
Rheumatoid disease - connective tissue	0.910	0.250	1.122	0.265
Peptic ulcer	1.096	0.557	1.015	0.929
Mild liver disease	0.876	0.218	1.375	0.002
Diabetes	1.066	0.125	0.997	0.950
Diabetes with complications	0.738	0.074	1.225	0.279
Hemiplegia	1.118	0.355	1.218	0.304
Moderate or severe renal disease	0.896	0.171	1.231	0.005
Moderate or severe liver disease	0.704	0.112	2.116	<0.001

Notes: CR = cost ratio, HR = hazard ratio, SIMD = Scottish Index of Multiple Deprivation. Grading system used = ICD-O/UICC with options 1: (Well) differentiated, 2: Moderately (well) differentiated, 3: Poorly differentiated, 4: Undifferentiated / anaplastic.

Multivariable regression results for breast cancer patients are shown in Table 3.8. A notable difference with trachea, bronchus and lung patients is that deprivation was positively associated with higher costs and with higher mortality.

Table 3.8.: Results of multivariable Cox regression on mortality risk and multivariable GLM regression on total eight-year post-diagnosis costs for breast cancer patients

	GLM		Cox	
Variable	**CR**	***p***	**HR**	***p***
Age				
< 50	1	Reference	1	Reference
50-59	0.837	<0.001	1.197	0.039
60-69	0.846	<0.001	1.853	<0.001
70-79	0.835	<0.001	3.510	<0.001
>= 80	0.673	<0.001	5.294	<0.001
Female	0.832	0.260	1.511	0.157
SIMD				
least deprived	1	Reference	1	Reference
2nd least deprived	0.956	0.157	0.989	0.862
3rd least deprived	1.006	0.862	0.990	0.883
2nd most deprived	1.032	0.309	1.176	0.013
most deprived	1.084	0.011	1.446	<0.001
Method of 1st detection				
clinical presentation	1	Reference	1	Reference
screening examination	0.844	<0.001	0.483	<0.001
incidental finding and other	0.906	0.055	1.164	0.095
Pre-diagnosis costs (per £1000)	1.016	<0.001	1.014	<0.001
Rural	1.036	0.116	0.994	0.894
Region network				
west	1	Reference	1	Reference
south and east	0.976	0.348	1.216	<0.001
north	0.842	<0.001	1.248	<0.001
Stage				
III	1	Reference	1	Reference
II or below	0.867	0.001	0.480	<0.001
IV	0.974	0.550	1.384	<0.001
HER				
negative	1	Reference	1	Reference
positive	1.243	<0.001	0.939	0.348
not known	1.017	0.493	1.080	0.140
Erstatus				
negative	1	Reference	1	Reference
positive	1.038	0.182	0.677	<0.001
not known	0.646	<0.001	0.977	0.847
Side				
right side	1	Reference	1	Reference
left side	0.988	0.541	1.003	0.951
not known	0.697	0.009	2.691	<0.001
Grade				
1 - low	1	Reference	1	Reference
2 - intermediate	1.175	<0.001	1.256	0.009
3 - high	1.407	<0.001	1.780	<0.001
not known	1.045	0.390	2.471	<0.001
Acute myocardial infarction	0.994	0.951	1.238	0.219
Congestive heart failure	0.882	0.314	1.417	0.009
Peripheral vascular disease	0.960	0.772	1.648	0.004
Cerebral vascular disease	1.020	0.821	0.942	0.730
Dementia	0.492	<0.001	1.345	0.060
Chronic pulmonary disease	1.118	0.035	1.190	0.066
Rheumatoid disease - connective tissue	1.047	0.647	1.211	0.407
Peptic ulcer	1.100	0.552	0.810	0.416
Mild liver disease	0.853	0.257	2.634	0.006
Diabetes	1.296	<0.001	1.496	<0.001
Diabetes with complications	1.386	0.147	1.363	0.250
Hemiplegia	0.959	0.845	1.134	0.776
Moderate or severe renal disease	1.190	0.155	1.383	0.062
Moderate or severe liver disease	0.649	0.095	6.843	0.016

Abbreviations: GLM = generalised linear model, CR = cost ratio, HR = hazard ratio, SIMD = Scottish Index of Multiple Deprivation, HER = human epidermal growth factor receptor

Results for colorectal cancer in Table 3.9 indicate that female sex lowered both costs (CR 0.937; p=0.002) and the hazard of death (HR 0.874; p<0.001). SIMD was not a significant factor for costs but raised the hazard of death significantly.

Table 3.9.: Results of multivariable Cox regression on mortality risk and multivariable GLM regression on total eight-year post-diagnosis costs for colorectal cancer patients

	GLM		Cox	
Variable	**CR**	***p***	**HR**	***p***
Age				
< 50	1	Reference	1	Reference
50-59	0.964	0.530	1.222	0.052
60-69	0.840	0.001	1.538	<0.001
70-79	0.758	<0.001	2.253	<0.001
>= 80	0.614	<0.001	3.606	<0.001
Female	0.937	0.002	0.874	<0.001
SIMD				
least deprived	1	Reference	1	Reference
2nd least deprived	0.968	0.330	1.135	0.019
3rd least deprived	0.973	0.397	1.151	0.009
2nd most deprived	0.990	0.752	1.217	<0.001
most deprived	1.044	0.207	1.279	<0.001
Method of 1st detection				
clinical presentation	1	Reference	1	Reference
screening examination	0.929	0.019	0.510	0.000
incidental finding and other	0.828	0.055	1.345	0.008
Pre-diagnosis costs (per £1000)	1.004	<0.001	1.011	<0.001
Rural	0.978	0.327	0.927	0.042
Region network				
west	1	Reference	1	Reference
south and east	0.956	0.079	0.994	0.887
north	0.957	0.111	1.001	0.990
Dukes' stage				
not known	1	Reference	1	Reference
A	1.381	<0.001	0.209	<0.001
B	1.535	<0.001	0.293	<0.001
C	1.630	<0.001	0.506	<0.001
C1	1.684	<0.001	0.445	<0.001
C2	1.672	<0.001	0.985	0.866
D	1.189	<0.001	2.188	<0.001
Grade				
1	1	Reference	1	Reference
2	1.067	0.269	1.103	0.383
3	0.941	0.332	1.582	<0.001
4	0.677	0.182	3.448	<0.001
not known	0.795	<0.001	1.622	<0.001
Acute myocardial infarction	0.998	0.983	1.150	0.181
Congestive heart failure	0.925	0.326	1.465	0.002
Peripheral vascular disease	1.055	0.463	1.059	0.615
Cerebral vascular disease	0.842	0.058	1.156	0.243
Dementia	0.603	<0.001	1.587	0.001
Chronic pulmonary disease	0.968	0.487	1.608	<0.001
Rheumatoid disease - connective tissue	1.133	0.313	1.493	0.004
Peptic ulcer	0.978	0.874	0.786	0.310
Mild liver disease	0.942	0.624	1.560	0.005
Diabetes	1.175	0.001	1.226	0.001
Diabetes with complications	1.115	0.560	2.073	0.014
Hemiplegia	1.404	0.442	1.301	0.424
Moderate or severe renal disease	0.982	0.827	1.333	0.005
Moderate or severe liver disease	0.528	0.034	4.449	<0.001

Notes: GLM = generalised linear model, CR = cost ratio, HR = hazard ratio, SIMD = Scottish Index of Multiple Deprivation. Grading system used = ICD-O/UICC with options 1: (Well) differentiated, 2: Moderately (well) differentiated, 3: Poorly differentiated, 4: Undifferentiated / anaplastic.

Results for patients with prostate cancer in Table 3.10 showed no significant associations between costs and age, and only at 70–79 years old (HR 2.34; p=0.006) and 80 years old and above (HR 3.644; p<0.001) were associations with hazard of death significant. While deprived areas were significantly associated with hazard of death, no significant associations were observed between SIMD quintile and costs.

Table 3.10.: Results of multivariable Cox regression on mortality risk and multivariable GLM regression on total eight-year post-diagnosis costs for for prostate cancer patients

	GLM		Cox	
Variable	**CR**	***p***	**HR**	***p***
Age				
< 50	1	Reference	1	Reference
50-59	0.963	0.728	1.073	0.825
60-69	1.034	0.751	1.344	0.339
70-79	1.145	0.193	2.344	0.006
>= 80	0.990	0.923	3.644	<0.001
SIMD				
least deprived	1	Reference	1	Reference
2nd least deprived	0.949	0.123	1.131	0.065
3rd least deprived	1.008	0.805	1.177	0.015
2nd most deprived	0.995	0.872	1.323	<0.001
most deprived	1.062	0.100	1.536	<0.001
Method of 1st detection				
clinical presentation	1	Reference	1	Reference
screening examination	0.934	0.214	1.283	0.015
Pre-diagnosis costs (per £1000)	1.014	<0.001	1.011	<0.001
Rural	0.961	0.108	0.895	0.020
Region network				
west	1	Reference	1	Reference
south and east	0.892	<0.001	0.944	0.287
north	0.923	0.005	0.993	0.904
Gleason score				
1	1	Reference	1	Reference
2	0.563	<0.001	0.000	<0.001
3	1.905	0.018	3.001	0.068
4	1.080	0.610	0.239	<0.001
5	1.052	0.746	0.389	0.015
6	0.612	0.050	0.420	0.061
7	0.850	<0.001	0.126	<0.001
8	0.979	0.539	0.203	<0.001
9	1.169	0.001	0.360	<0.001
10	1.174	<0.001	0.540	<0.001
not known	1.339	0.001	0.947	0.720
Acute myocardial infarction	1.065	0.511	0.993	0.964
Congestive heart failure	0.801	0.048	1.305	0.157
Peripheral vascular disease	0.900	0.226	1.117	0.498
Cerebral vascular disease	1.144	0.118	0.879	0.384
Dementia	0.407	<0.001	2.731	<0.001
Chronic pulmonary disease	1.114	0.092	1.364	0.006
Rheumatoid disease - connective tissue	1.065	0.662	1.090	0.746
Peptic ulcer	0.935	0.810	0.877	0.721
Mild liver disease	1.095	0.530	1.250	0.499
Diabetes	1.117	0.068	1.168	0.183
Diabetes with complications	1.204	0.250	0.883	0.603
Hemiplegia	1.210	0.312	2.143	0.058
Moderate or severe renal disease	1.002	0.984	1.377	0.026
Moderate or severe liver disease	0.691	0.009	3.992	<0.001

Abbreviations: GLM = generalised linear model, CR = cost ratio, HR = hazard ratio, SIMD = Scottish Index of Multiple Deprivation,

3.4 Discussion

3.4.1 Key Results

Resource use and associated costs were substantial, as seen in other studies, with markedly higher resource use in the year after diagnosis than in other years. Considerable variation between cancers was observed. The highest eight-year costs were found in non-Hodgkin lymphoma and the lowest in malignant melanoma of skin. The highest rates of cost accrual were observed in the initial and end-of-life phases, but cumulative costs over the eight-year follow-up were highest in the continuing phase. A complex relationship between survival and costs was observed across cancer types. Malignant melanoma of skin had the highest survival rate but also the lowest cumulative costs, while cancers with very low survival such as trachea, bronchus and lung had lower cumulative costs than cancers with higher survival such as non-Hodgkin lymphoma. Factors positively associated with higher costs tended to be negatively associated with mortality. Exceptions were pre-diagnosis costs and diabetes, both having positive associations with costs and mortality, and screening as method of first detection, which was negatively associated with both costs and mortality. Costs for all cancers combined showed significant negative associations with being female (CR 0.942; $p<0.001$), which were stronger in multivariable models (CR 0.819; $p<0.001$). Being female was also associated with lower hazard of death in univariable models (HR 0.962; $p<0.001$) and multivariable models (HR 0.975; $p=0.022$). However, for lung cancer, being female had a positive association with costs (CR 1.086; $p<0.001$) while retaining a negative association with hazard of death (CR 0.886; $p<0.001$). Resource use was considerably higher in the year after diagnosis than in other years for all cancers. However, there was high variation between cancers in the magnitude of the year 1 costs.

3.4.2 Interpretation

Risk Factors

More advanced stage had a strong positive association with hazard of death, however the associations with costs were more complex. Both early-stage and late-stage cancers showed negative associations with costs, with stage IV cancers being associated with the lowest costs. Comorbidities showed generally positive associations with hazard of death. Where negative associations existed they were not significant. Associations with costs varied, with dementia showing a strong negative association

with costs (CR 0.513; $p<0.001$) in multivariable models for all cancers combined, while diabetes showed a positive association (CR 1.103; $p<0.001$) and other conditions varied. In general, factors positively associated with hazard of death tended to be negatively associated with costs, reflecting the importance of survival on long-term costs. An exception was that diabetes was associated with both higher costs and higher mortality. The association with stage was non-linear, with the highest costs tending to be associated with mid-stage cancers. As different cancers tend to be detected at different stages this could explain some of the variation between cancer types. Why mid-stage cancers have higher costs could be a result of the probability of curative treatment being given and the treatment's aggressiveness. Patients with early-stage cancers are more likely to survive longer, but will require less aggressive treatments. Patients with late-stage cancers are more likely to die sooner and less likely to undergo curative treatment, with given treatments likely to be less aggressive, which would result in lower costs.

Regional Costs

Reference costs in the Scottish Costs Book showed variation in costs across different regions and between hospitals within regions. Capturing regional information about costs would have involved considerable complexity hence was not performed. The use of HRGs is an established method of costing healthcare use [108], however these are not used in Scotland and methods based on English groups may not have captured variation within Scotland and would also lack regional costs specific to Scotland, which are likely to vary due to the remoteness of some Scottish regions. The lack of regional costings in the Scottish Costs Book should not have caused bias if the averages in the Scottish Costs Book used appropriate weightings, however information on the weightings used was not given. If the costs of remote regions, which are likely to have higher costs, were given higher weightings relative to their populations than other regions, this have biased costs upward. Cross-border healthcare (i.e., healthcare performed beyond Scotland) was not recorded, which could have led to the underestimation of costs.

Variation Across Cancer Types

A striking feature of the results was the high variation between cancer types. However, the reference costs used were not cancer-specific, meaning that the costs of treatments and drugs for specific cancers were aggregated across all cancers. Lengths of inpatient stays were also not accounted for. It is therefore possible that the variation in costs

between cancer types was underestimated (while counts of resource use and monetary costs of prescription would be unaffected). As costs were unadjusted, differences in patient factors between cancer types seem likely to have contributed to variation in costs between cancers. However, some associations between patient factors and costs varied in direction across cancer types. Higher deprivation was associated with higher hazard of death, but no clear pattern was observed with costs over all cancers combined. For trachea, bronchus and lung cancers, higher deprivation was significantly associated with lower costs and higher hazard of death, whereas for breast higher deprivation was significantly associated with higher costs and higher hazard of death. Other factors showed more consistent associations. Pre-diagnosis costs showed a significant positive association with higher costs and with higher hazard of death in all cancer types analysed and in all cancers combined. It is not clear why this should be, as it contrasted with the general pattern of more comorbidity being associated with lower costs. Serial dependence, where people who accrue high costs before diagnosis continue to do so after diagnosis, may be an explanation. However, this gives little insight into what drives the higher costs. Rurality varied across sites; for all cancers combined there was a negative association with costs and with hazard of death. This was also the case with colorectal cancer, while respiratory cancers had a negative association with costs and a positive association with hazard of death and associations for breast were not significant. Screening was significantly associated with lower costs and lower hazard of death, suggesting that improved screening could improve survival rates while reducing costs. However, the method of first completion variable had many missing values suggesting a cautious interpretation of the associations.

Variation Over Time

In all phases of care, inpatient costs contributed more to total costs than outpatient and prescription costs combined. However, their contribution was lower in the continuing period, reflecting lower hospital stays and higher outpatient visits and prescription costs. Prescription costs were comparable to outpatient costs and became relatively more important over the long-term, making a substantial contribution to overall costs in the continuing period. Overall, monthly costs were highest in the end-of-life phase and lowest in the continuing phase for all cancers. However, the cumulative effect over the entire follow-up meant that for all cancers but lung cancer and oesophagus cancer, the continuing phase had higher costs than the initial phase, and for all cancers except malignant skin melanoma, non-Hodgkin lymphoma and oesophagus cancer, the continuing phase had higher costs than the end-of-life phase.

Cost Assignment

Inpatient costs have been found to be the largest proportion of direct costs in other studies [172]. This analysis found that inpatient costs remained the largest component of costs throughout an eight-year follow-up. However, the method of calculating inpatient costs can substantially affect their magnitude. Furthermore, when costs are measured over a longer period, discrepancies between costing methods may be amplified by the presence of longer stays. PLICS may become the de facto standard for costing healthcare use but only had only good coverage for years 2015 onwards in the SMR datasets available to me. While costings based on HRGs may be more accurate, per-episode approaches are considered to be of acceptable accuracy [108]. As the foci of my analysis were the dynamics of costs and differences between cancers, rather than precise costing, per-episode costings gave an acceptable compromise between precision and complexity. Furthermore, HRGs based on English cost groupings may not be appropriate for healthcare in NHS Scotland. Healthcare costs tend not to be normally distributed and are often dominated by large numbers of zero values, requiring the use of models such as GLM and two-part regression [134]. Taking costs over longer time periods and including ongoing costs like prescriptions reduced the number of zero values in my analysis, which may have improved statistical properties and made modelling more accurate. However, processing such a large number of prescription records incurred a high computational cost, while prescriptions were a minor proportion of total costs.

Comparison With Other Studies

Details of the studies discussed in this section are given in Table 4.1. Banegas et al. (2018) [135], using a phase-of-care approach for total and net costs, found lung cancer more expensive than breast, colorectal and prostate cancers, in contrast to my findings. However, breast and colorectal were found to have similar costs and prostate lower, in agreement with my measurements. Costs overall were considerably higher than my measurements—converted to sterling—but this study measured a US population of patients on a health plan, with costing methods geared to the US private healthcare system. Higher drug costs in the US may partly explain the higher costs overall, and population differences may also have contributed. This may also have been the case with Yabroff et al. (2008) [21] who like myself studied a wide range of common cancers and found costs highest in the initial and end-of-life phases, with survival and stage as major cost factors. Other areas of agreement were the high costs incurred for non-Hodgkin lymphoma, head and neck cancer and oesophagus

cancer, and much lower costs for skin cancer and prostate cancer. However, as with Banegas et al. (2018) [135] and unlike my results, lung-related cancers were found to have relatively high costs. The large sample of US Medicare patients may be more comparable to my sample than that of Banegas et al. (2018) [135], but was limited to patients aged 65 and over and only covered fee-for-service patients—around 85% of Medicare patients—so may have suffered from selection bias. The high cost of lung cancer in US studies may be due to differentials in healthcare systems between the UK and the US, where healthcare costs can be considerably higher.

UK Studies

Studies of UK populations may provide closer comparisons to my analysis. Laudicella et al. (2016) [105] calculated nine-year incidence costs for NHS England cancer survivors, with results more similar in magnitude to my findings than those of Banegas et al. (2018) [135] and Yabroff et al. (2008) [21]. However, the method measured costs only for patients alive at the start of each follow-up period, with the means in each period summed to provide nine-year cumulative costs. While this method is likely to be informative for costs accruing to survivors, it contains a form of survivor bias that is likely to overestimate societal costs; by considering only costs for cancer survivors, it ignores the economic costs accruing to mortality, as the counterfactual—the costs that would have occurred if the patient had not died of cancer—is unmeasured. Marti et al. (2015) [146] also found breast and colorectal more expensive than prostate and that a small number of patients incurred very high costs, however, the small sample size, different population and focus on societal costs, including out-of-pocket costs, in addition to the shorter follow-up period may make results less comparable. Hall et al. (2015) [63] using HRG and PLICS costings to measure 15-month cumulative costs for patients in England, also found breast and prostate to have similar costs that were notably higher than prostate, and that stage was a strong predictor of costs. However, this study used the same ePOCS study data source as Marti et al. (2015) [146] with relatively low numbers so similar problems of comparability may apply.

Stage

End-of-life care costs have been observed as generally higher for later-stage cancers [135]. However, one other study found localized tumours to be more expensive than metastatic tumours and regional ones to be most expensive of all [173], in line with my results. Additionally, for later-stage cancers the end-of-life costs will occur sooner and for cancers with very high mortality at advanced stage, such as lung cancer, the majority of end-of-life costs will be recorded in the study period. For lower-stage cancers, more patients will undergo curative therapy and the proportion being recorded

during the end-of-life phase will be lower in studies of limited duration. An advantage of the phase-of-care approach is that it should reduce this discrepancy, however an analysis using the phase-of-care approach still reported considerably higher overall costs for later-stage cancers [135].

Other Risk Factors

Previous studies have found sex, type of cancer and socioeconomic factors to be significant determinants of healthcare resource use [67, 66]. Age has also been reported as significant but the direction of association varies between and even within studies [66]. My results showed variation between cancers in the significance and direction of associations with costs for sex and SIMD, with no clear pattern visible. The general direction of the association between age and costs was negative, though a non-linear relationship was observed in prostate cancer. Magnitudes of associations and statistical significance were unexpectedly low for SIMD. A possible explanation is that the effect of deprivation status was linked to the cancer type. This may have caused associations to be negated when analysing all cancers combined, as associations in different cancer types took opposite directions.

3.4.3 Strengths

Patient-level data covering the Scottish population provided a large and varied enough sample to describe resource use for the 10 most common cancers and other cancers combined. Other studies have tended to focus on four common cancers: lung, breast, colorectal and prostate. An eight-year follow-up provided information on how resource use and associated costs developed over time. By reporting units of yearly resource use this study allows other researchers to assign custom unit costs, which may be specific to a health system or to particular cancers. Reporting of yearly costs provided information on how costs changed over time at the cohort level, while also allowing other researchers to calculate cumulative costs with custom discount rates.
Phase-of-care trajectories provided information on how costs change for survivors and can be adapted by other researchers to calculate long-term costs with custom mortality rates and perform sensitivity analysis. The inclusion of community prescriptions improved the statistical properties of cost data, reducing the problem of low zero counts commonly present in healthcare costs. It also provided information about ongoing resource use long after diagnosis, giving a fuller account of long-term costs. Comparing predictors of survival with predictors of costs provides information on how these are related, while giving insights into the factors driving costs and mortality in the Scottish population.

3.4.4 Limitations

Datasets

There were also limitations. Public administration data are believed to be a weak source of information on comorbidities [118] and the SMR data have known limitations in accuracy [161]. Common conditions are under-recorded in SMR01 [161] meaning that confounding effects in regression models may not be fully accounted for which could bias the coefficients of other explanatory variables. An issue with outpatient data was that in previous years only new consultations were compulsory to report for some health boards, which may have led to under-reporting of outpatient visits. This was likely to have caused underestimation of costs, however the low contribution of outpatient costs makes it unlikely this was substantial. Additionally, geriatric long-stays were excluded from SMR01 records until 2007 which may have led to underestimations of inpatient episodes.

Variables

TNM stage was only recorded for lung and breast cancer. The complex stage data in less common cancers were challenging for a non-clinician to interpret, therefore a degree of simplification was required in models that analysed all patients rather than particular cancers. While this may have led to some bias in the coefficients of stage and other variables, my findings generally concurred with other studies that found stage an important predictor of costs [63, 105]. Additionally, the datasets lacked useful variables describing frailty, adiposity, ethnicity and marital status, which may have led to biased associations. The lack of precise dates in the prescribed items data meant prescription costs could only be accurately assigned to the nearest year, and that some prescription costs would be assigned to incorrect time periods. It is not known whether this would introduce bias and the relatively small contribution of prescription costs suggests that any such bias would be limited.

Outcomes Data

Data on people who moved abroad were not available in my dataset. As these individuals would still be counted in the follow-up and would accrue zero costs after emigrating (even if they used healthcare abroad), costs would have been underestimated. Complementary healthcare was not recorded, which could also bias cost estimates downward. Another issue was that around 8.5% of the Scottish

population use some form of private healthcare [174] which may also have exerted downward bias on cost estimates due to non-capture of healthcare use data. However, it could be argued that lack of private healthcare data better represents the actual outlays by NHS Scotland.

In addition to these factors, a lack of information on healthcare outcomes is likely to have caused underestimations of total healthcare costs. Cancer patients have been reported to accrue higher expenditure relating to mental health [175]. The inclusion of SMR04 records, which relate to mental health, would have given a more complete picture of healthcare expenditure. The possibility of accessing this dataset was examined but the additional ethical and privacy issues burdens were considered too high. The lack of information on mental health conditions seems likely to have biased costs downward, and may have led to underestimations of variation between cancer types. Social care is also a part of Scotland's NHS hence a full account of healthcare resource use would include social care costs. Data on primary care were also unavailable. A wider estimate of costs to society would include those accruing to lost productivity and out of pocket costs such as parking charges, however such costs were beyond the scope of this study. Costs of prevention, education and information relating to cancer awareness were also not under investigation, though they could legitimately be included as healthcare costs related to the disease.

3.4.5 Generalisability

The health of Scotland's population is known to differ from other European nations [56], and its geography and healthcare system are also unique, hence generalisability may be limited. Even within the UK there are differences that may limit the external validity of results. Prescriptions are free for patients in Scotland, which may increase demand relative to countries where patients must pay the full price or a contribution to the price, as in England. Hence prescription costs may not generalise to other populations. However, prescription costs were a relatively small component of costs. The remoteness of much of Scotland should be taken into account as remote communities have been found to incur higher health costs [82]. Units of healthcare resource use are likely to be more generalisable than monetary costs due to uncertainty in unit costing, differing costs of healthcare, and different healthcare systems. Despite the differences noted above, it is likely results will generalise to European nations better than the US, due to greater similarity to European healthcare systems.

3.4.6 Policy Implications and Future Research

Cancer Types

People with the four most common cancers in Scotland (lung, breast, colorectal, prostate) accounted for almost half of all the spending on healthcare measured in this analysis. However, some less common cancers, most notably non-Hodgkin lymphoma, saw higher per-person spending. Combined, the spending on patients with less common cancers was higher than the four most common cancers combined. This, and the considerable cost variation between cancers, suggest notable cost reductions could be found by targeting less common cancers with high costs such as non-Hodgkin lymphoma and head and neck cancer.

Survival and Costs

While a complex relationship between survival and costs across cancer types was observed, at the patient level, factors associated with higher survival tended to be associated with higher costs. If cancer survival improves, policymakers should anticipate higher healthcare costs associated with cancer. However, while an ageing society may increase cancer incidence, higher age had a negative association with costs, which may counterbalance the additional numbers of cancers. Furthermore, cancers with low survival tended to have low costs, while cancers with moderate survival (and some with high survival) had higher costs. If cancers with low survival fall in incidence while more costly cancers rise, the costs to healthcare systems and wider society are likely to rise. Hence policymakers may be presented with the problem that reducing deaths from cancer will result in increased economic costs. A possible solution suggested by this analysis is early detection. However, if cancers are detected at regional rather than distant spread, the costs may be increased due to the extra healthcare required during treatment and afterwards. My results indicated that early stage detection can reduce post-diagnosis healthcare, suggesting that screening may be more cost effective if targeted at younger people. But against this must be balanced the economic costs of screening and other burdens such as the psychological effects of screening and false positives.

Long-Term Costs

While the rates of cost accumulation were highest in the treatment and end-of-life phases, over the long term costs in the continuing phase were the most substantial in magnitude for patients who entered this phase. The contribution of prescriptions

became more pronounced during the continuing phase and was of similar magnitude to outpatient visits. This suggests that the omission of prescriptions from the long-term costs of disease may bias the costs downwards, but also suggests an area for further investigation where costs might be reduced.

Data

Although Scotland has considerable amounts of clinical data available to researchers, improvements could be made in the reporting of cancer data, particularly for tumour stage and for presence of particular comorbidities. The increased adoption of PLICS could simplify costing studies and improve their accuracy and comparability. Enhanced reporting of socioeconomic variables would also be of value to researchers.

3.4.7 Conclusion

In this analysis I linked NHS healthcare datasets to measure patient-level healthcare use and associated costs of cancer patients over eight years in the Scottish population. The large sample size allowed me to break down results by the 10 most common cancers and other cancers combined. Incorporating prescriptions provided a fuller account of healthcare costs over the long term. Comparing risk factors for survival to risk factors for costs provided new understanding of how survival and costs are related. This knowledge will help healthcare professionals, health economists and policymakers allocate resources efficiently and better understand the long-term outcomes of cancer. But questions remain. How do these costs compare to a similar control group without cancer? How does existing morbidity affect the costs? What are the effects of cancer on employment? In the chapters that follow I will attempt to answer these questions using additional analyses.

4 Comparison with a Matched Control Group

4.1 Introduction

4.1.1 Background

The findings described in Chapter 3 provided novel information on the economic costs of cancer and how they changed over time in the Scottish population. Comparison of cancer types and analysis of risk factors provided insight into what drove costs, however, studying cancer patients in isolation has limitations. Natural ageing is likely to have influenced the cost trajectories [111], making the long-term impact of cancer on costs challenging to measure without a suitable comparison. High costs were observed in the end-of-life period, however, this phase is also associated with high costs in patients with other conditions [176, 55]. The high mortality rates of certain cancers, some of which have one-year survival rates below 50% [22], imply that a notable proportion of patients will enter the end-of-life phase during the year after diagnosis. As everyone must ultimately die, it may be asked to what extent costs associated with cancer are end-of-life costs brought forward, rather than excess costs above the population baseline. In addition to ageing, other factors may have affected the cost trajectories, such as improvements in treatment and wider economic factors such as the 2007–8 Global Financial Crisis (GFC) and subsequent government austerity. A control group without cancer would help to disentangle costs arising from cancer from those caused by patient factors like age and by other factors.

Excess costs, also known as net costs, can help to isolate the impact of cancer on costs. They are defined as the difference between the mean costs of cancer patients and those of similar patients without cancer [96]. Measuring excess costs allows researchers to separate changes in costs over time that are caused by cancer from normal trends such as ageing, and secular factors like changes in healthcare systems and treatments [96]. An alternative method of calculating the costs associated with cancer is to categorise each unit of resource as being cancer-related or not, and take the mean of the costs attributed to cancer. Computing attributable costs in this manner is more complicated and error-prone because attributing costs to a particular disease is not straightforward [96]. Furthermore, attributable costs cannot be negative, meaning that cancer costs could be overestimated if non-cancer patients use more resources overall. As calculating excess costs is simpler, more transparent and less prone to bias, it is generally preferred [96].

Substantial excess costs have been found in multiple studies of cancers of the lung [105, 177, 178, 135], breast [105, 178, 179, 21, 135, 180, 181, 182], oesophagus [21]

and for colorectal cancers [105, 178, 179, 135, 183, 184, 21]. Lower but still substantial excess costs have been found for cancers of the prostate [105, 178, 179, 135, 185, 186] and skin [179, 21]. These results and others discussed are summarised in Table 4.1 and were identified using the strategy given in Section 3.1 but filtered by costs that were described as net or excess with the study including a comparison group without cancer. Due to their high prevalence these cancers contribute heavily to population-level costs [105, 21], but less common cancers may be more costly at the patient level [179]. To compare costs across countries and time, I shall, as in Chapter 3, convert costs to sterling using PPP rates from the OECD [149] and inflate them to 2018 price levels using rates from the BoE [150]. A 2015 study in the New Zealand (NZ) population study found the highest excess costs in bone and connective tissue cancers at £52,793 compared to £15,623 for lung cancer, £24,241 for breast cancer, £8,619 for prostate cancer, £23,164 for colorectal cancers and £4,310 for skin cancers [179]. Studies from the US and the UK that compared multiple cancer types have also found considerable cost variation between sites [21, 135, 105, 186] and substantial excess costs have been observed for all cancers combined [179]. Higher excess costs than those given by Blakely et al. (2015) [179] are found in US studies. For instance, Yabroff et al. (2008) [21] observed five-year excess costs of lung cancer at £33,977 for men and £35,503 for women, prostate cancer at £18,450 and colorectal cancer at £36,621 for men, £35,037 for women, however female breast cancer costs were relatively low at £17,001. In a 2018 US study Banegas et al. (2018) [135] found even higher five-year excess costs of £61,053 for lung cancer, £40,541 for breast cancer, £18,570 for prostate cancer, and £37,283 for colorectal cancer [135]. Excess costs for cancer survivors have been found to stay elevated compared to similar controls at up to ten years after diagnosis for breast cancer [181]. In terms of populations and health systems, studies of more relevance to the Scottish population are likely to be Blakely et al. (2015) [179] and Laudicella et al. (2016) [105], which measure excess costs for the New Zealand and English populations respectively. However, Laudicella et al. (2016) [105] estimated excess costs as prevalence costs over the entire population rather than patient-level incidence costs, which gives little information about how costs develop over time. Furthermore, the English NHS and Scottish NHS are distinct organisations, and transposing unit-costing methods may give biased estimates [108]. This consideration is also likely to limit the generalisation of the results of Blakely et al. (2015) [179]. Scotland's population has poor health outcomes as discussed in Section 4.1.1, which may further limit generalisation of results, even with studies carried out in the English population.

Table 4.1.: Summary of literature results of the excess costs of cancer

Study	Year	Currency	Population	Methods	Main results
Laudicella et al.	2016	UK sterling	NHS England patients	9 year incidence + 5 year prevalence	Incidence costs in the first year of diagnosis noticeably higher in patients age 18-64 than age >= 65 across all examined cancers. A lower stage diagnosis is associated with larger cost savings for colorectal and breast cancer.
Banegas et al.	2018	US dollars (USD)	45,522 cancer and 314,887 controls using us health plan data	Total and net costs using phase of care approach. 1 year and 5 year costs reported	Net costs were consistently highest for lung cancer and lowest for prostate cancer. Net costs were higher across all cancer sites for patients aged <65 years than those aged ≥65 years. Medical care costs for all cancers increased with advanced stage at diagnosis.
Yabroff et al.	2008	USD	718,907 cancer and 1,623,651 non-cancer us medicare patients (65+)	Net costs by phase of care using survival data to give 5 year costs	Mean net costs of care were highest in the initial and last year of life phases and lowest in the continuing phase. Mean 5-year net costs varied widely, from less than $20 000 for patients with breast cancer or melanoma of the skin to more than $40 000 for patients with brain or other nervous system, esophageal, gastric, or ovarian cancers or lymphoma.
Kutikova et al.	2005	USD	2040 lung cancer patients who were employees of large corporations in the US	Retrospective case control, inpatient, outpatient and drug costs, 2 years from diagnosis	Regression-adjusted mean monthly total costs were US dollar 6520 for patients versus US dollar 339 for controls (P < 0.0001), and overall costs across the study period (from diagnosis to death or maximum of 2 years) were US dollar 45,897 for patients and US dollar 2907 for controls (P < 0.0001). The main cost drivers were hospitalization (49.0% of costs) and outpatient office visits (35.2% of costs).
Pisu et al.	2018	USD	Multiple	Review of studies	For all payers combined, costs for cancers like breast, prostate, colorectal, and lung cancers were $20,000 to $100,000 in the initial phase, $1000 to $30,000 annually in the continuing phase, and >=$60,000 in the end-of-life phase.
Blakely et al.	2015	NZ dollars	New zealand cancer registry	Gamma regression on expected costs	From $5,000 (melanoma) to $66,000 (bone and connective tissue).
Lang et al.	2014	USD	Seer medicare stage IV breast cancer patients	Matched 1:1 retrospective analysis	Higher resource use in all areas except oral prescriptions.
Hanchate et al.	2010	USD	452 cancer patients and, 1,656 matched controls from various locations in the US	Prospective five-year longitudinal comparison of cases and matched controls	Breast cancer survivors' health care use and disease burden return to pre-diagnosis levels after one year but greater use of outpatient care persists at least five years.
Khanna et al.	2011	USD	West Virginia Medicaid administrative claims data for women recipients 21-64 years of age	Matched controls analysis	All-cause healthcare costs significantly higher for breast cancer patients than controls ($16,345 vs. $13,027, p<0.001).
Song et al.	2011	USD	6,675 patients with colorectal cancer matched to patients without cancer	Retrospective study using national claims database	Total monthly costs were $14,585, driven by higher inpatient care ($7,546) and outpatient care ($6,749).
Chang et al.	2004	USD	New diagnoses of one of seven types of cancer (n = 12,709). Controls without cancer were matched at a 3:1 ratio by demographics	Retrospective matched-cohort control analysis	Mean monthly costs ranged from 2,187 dollars for prostate cancer to 7,616 dollars for pancreatic cancer, most often driven by hospitalization. Costs for controls were 329 dollars per month.
Jayadevappa et al.	2005	USD	120 prostate cancer patients and 240 men without cancer, matched by age and race	Retrospective cohort control study using regression models	The incremental cost of prostate cancer was 1.30 times higher than controls.

This evidence indicates that cancer incurs substantial excess costs, which vary by site and stage. However, evidence is lacking in the Scottish population, and costs for other populations with different healthcare systems may not generalise for reasons discussed in Section 3.4. Additionally, while cancers with high prevalence are likely to comprise a large proportion of overall cancer costs in a society, the combined impact of less common cancers may also be substantial but these have received little study in the UK. This analysis aimed to address the literature gaps by comparing the healthcare use and associated costs of cancer patients to similar individuals without cancer in the Scottish population, and estimating long-term excess costs.

4.1.2 Aims and Objectives

To address gaps in the existing literature and provide cost estimates for cancer costs in the Scottish population, I performed an analysis using the cancer cohort described in Chapter 3 and a matched control group of individuals without any history of cancer. The overall aim was to better understand how the healthcare use and associated costs of a cancer cohort compared to a similar cohort without cancer, with emphasis on trends in costs over time. Specific objectives were as follows.

1. Measure the long-term excess costs of cancer in the Scottish population.
2. Chart and compare trajectories of resource use and associated costs over time for people with and without cancer.
3. Determine when, if at all, mean excess costs reverted to zero after the cancer diagnosis.

4.2 Methods

4.2.1 Study Overview

This was a retrospective matched pairs cohort study, using individual-level data from linked administrative datasets. A total of 55,807 patients with cancer and 55,807 patients without cancer were followed over eight years. As in Chapter 3, the incidence approach was used to measure direct costs from a healthcare-payer perspective. The setting and study periods were identical to those described in Chapter 3, covering all regions of NHS Scotland with exposure during the period 1 January 2009 to 31 December 2010 and a subsequent follow-up of eight years. GLM regression methods were used to estimate excess costs adjusted for additional confounders not covered by matching.

The exposed cohort was identical to that in Chapter 3, with exposure defined as a first cancer diagnosis recorded in SMR06 during the entry period, and cancer types derived from the same ICD10 codes. Other eligibility criteria for the cancer cohort were identical to those described in Chapter 3. Each cancer patient was matched with a single unique individual with no history of cancer, using exact matching on the year of birth, sex, SIMD quintile, and NHS Scotland Health Board of residence at the time of the cancer patient's diagnosis. Considerations regarding the study size described in Section 3.2 also applied to this analysis. As the analysis used 1:1 matching for all cancer individuals, this gave a total sample size of 55,807 x 2 = 111,614 unique individuals.

4.2.2 Data

The cancer cohort was extracted from SMR06 as described in Chapter 3. Each individual in the cancer cohort was matched with an individual without cancer, but with similar characteristics, listed in the CHI database. Due to the sensitive nature of such data, researchers were not given access to raw datasets, consequently, matching

was performed by eDRIS in a secure environment. The high level of detail in these datasets allowed exact matching on demographic data, however data on health was limited. Health outcomes for the cancer and control cohorts were extracted from SMR00, SMR01, and PIS as described in Section 3.2.3, using CHI numbers to link the separate datasets. Dates of death for the cancer cohort were extracted from SMR06, however this was not possible for the non-cancer cohort, hence a further linkage to National Records of Scotland (NRS) Death Records dataset was carried out. Linkage and anonymisation of the cohort datasets were carried out by eDRIS in a secure environment inaccessible to researchers. The outcome datasets: SMR00, SMR01, NRS Deaths and PIS, were linked to the cohort dataset by myself in the National Safe Haven, which is a secure environment where researchers can analyse data.

4.2.3 Variables

Outcome Variables

The primary outcomes were total and excess costs. Calculation of excess costs followed methods that will be described in Section 4.2.4. Other outcomes were units of resource use (inpatient episodes, inpatient days, outpatient visits, prescribed items), and survival times. Units of resource use were identical to those described in Section 3.2.4 and calculated using the same methods for both cancer and non-cancer cohorts, as were the associated total costs. The method of calculating survival times differed slightly from that of Chapter 3, because date of death was not available in the CHI dataset and hence had to be extracted from the NRS Death dataset. However, the NRS records were precise only to the nearest month. Date of death for the cancer cohort was extracted from SMR06 records as described in Section 3.2.5, and rounded to the nearest month to provide equivalence with the non-cancer cohort.

In Section 4.2.4 and Section 4.4 categories of low, medium and high survival are mentioned. These categories were for discussion of results only and were not defined in the dataset itself. I followed the boundaries given by Blakely (2015) [179] for five-year survival with low < 0.25, moderate $= 0.25$–0.6, high > 0.6. While Blakely's study used five-year survival the difference between five-year and eight-year survival in my dataset was minimal and made no practical difference to the cancer sets encompassed.

Exposure Variables

Inclusion in the cancer cohort was indicated by a simple binary variable. Additionally, each individual record contained an identifier link to its pair in the matched cohort,

enabling stratification by cancer site in both the cancer and non-cancer cohorts; the cancer site for a cancer-free patient was that of his or her matched pair. To follow individuals over time, each individual was given a diagnosis-event date, which for cancer patients was the date of the first cancer diagnosis and for cancer-free individuals was a pseudo-diagnosis at an equivalent date, relative to the person's date of birth. Hence if a cancer patient received a diagnosis on her 65th birthday, a diagnosis-event would be created on the 65th birthday of his or her matched pair in the non-cancer cohort. Phases of care were created for both cohorts relative to the diagnosis-event, as described in Section 3.2. Patients without cancer were assigned to the treatment phase in the year after the diagnosis-event similarly to cancer patients, however this did not imply that any treatment was given. The continuing phase and end-of-life phases were assigned using the same criteria as cancer patients as described in Section 3.2.

Confounders

Individuals were precisely matched on year of age, sex, SIMD quintile and NHS Health Board, therefore no further adjustment was performed on these variables. As the SIMD quintiles and Health Boards were geographically broad, the rurality variable described in Section 3.2.5 was used for further adjustment. Comorbidities and pre-diagnosis costs were extracted for both cohorts using the methods described in Section 3.2.4. As in Chapter 3, if no pre-diagnosis records existed for an individual, it was assumed that the individual was free of all listed conditions and had zero pre-diagnosis costs. Although the complete list of comorbidities used in Chapter 3 was present in the dataset, not all comorbidities were found to be significant risk factors. While all comorbidities went through the variable selection process described in Section 3.2.5, only 10 were ultimately used in the final regression models. As numbers for the other variables tended to be very low, they were not included in the baseline description, as this could have caused issues with disclosure. To further adjust for residual confounding due to additional comorbidities, a comorbidity count was also included. Model selection based on sums of squared residuals indicated a categorical variable with three levels (0, 1, 2 or more comorbidities) gave a better fit compared to a simple count, Charlson score or a categorical variable with additional levels. Pre-diagnosis costs were not included in regression models, although these were shown in Chapter 3 to be risk factors. The trajectories indicated an increase in excess costs during the year before diagnosis, hence inclusion of pre-diagnosis costs may have introduced bias through endogenous effects.

4.2.4 Statistical Methods

To understand how survival over time varied across cohorts, I plotted Kaplan-Meier (KM) survival charts, stratified by cancer type. Further trajectories described total costs and units of resources used: inpatient/daycase episodes and days, outpatient visits and numbers of prescribed items. Separate charts were produced to show stratifications by year relative to diagnosis-event and by phase-of-care as described in Section 3.2.6.

To estimate eight-year excess costs I used univariable and multivariable GLM regressions. The rationale for using GLM models and the assumptions underpinning them are described in Section 3.2.6, but while the distributional considerations were common to both analyses, the purpose of using regression modelling diverged. In Chapter 3 the purpose was to understand the factors contributing to costs and survival, while in this analysis the aim was to estimate the excess costs of cancer, adjusted for confounding factors. The matched data in this analysis necessitated different methods of dealing with standard errors. Matched data cause clusters to form across pairs of individuals, invalidating regular standard errors. Alternative regression methods exist to account for matched data [187], but these were not appropriate for the continuous nature of cost outcomes. Instead, I specified robust standard errors that accounted for clustering. For the GLM specification I again used the gamma distribution with log link, the maximum likelihood method to estimate coefficients, and the eform option to describe each coefficient as a cost ratio (CR), as described in Section 3.2.6. Average marginal effects on the cancer variable produced the excess cost estimates and 95% confidence intervals, and model fit was assessed using sums of squared residuals and visual inspection of Anscombe residuals and quantile-quantile (Q-Q) plots. Outputting residuals charts from the Safe Haven is forbidden because they depict single data points, and therefore only the approximate shapes could be described. These tended to deviate from normality due to the skewed nature of health data, but gave no indication of bias.

The univariable model took specification

$$y_i = \beta_0 + \beta_1 cancer_i + \varepsilon_i \tag{4.1}$$

where y_i represents the eight-year aggregate of total costs for an individual, $cancer_i$ is a binary variable representing whether an individual i has cancer, α measures the

baseline costs at the baseline value of $cancer$, which are equal to the costs for the comparison cohort, ε_i is a random error term, and β_1 represents the excess cost of cancer. The link function is omitted for clarity. Therefore, the univariable model was implicitly adjusted for matched variables age, sex, SIMD quintile and NHS Health Board. An individual regression was run for each cancer site. To adjust for additional confounding variables, I ran multivariable regression analyses with the base model

$$y_i = \beta_0 + \beta_1 cancer_i + \beta_x X_i + \varepsilon_i \tag{4.2}$$

Where y_i, β_0, and ε are per the univariable model, β_1 is again the excess cost of cancer, and X_i is a vector of independently distributed variables: rurality, individual comorbidities, and a further categorical comorbidity count described in Section 4.2.3, with β_x the variable's coefficient. Again, individual regressions were performed for each cancer type under study. As model parsimony was not a priority, and the coefficients of confounders were not under study, a non-parsimonious approach to variable selection was taken. Confounding variables not already covered by matching were included based on the literature and the risk factors identified in Section 3.3.5.

While charts of costs and resource use were appropriate for showing the general shapes of trajectories, the wide ranges of costs over time entailed sacrificing precision in order to show the full eight-year follow-up at a meaningful scale. To test the equality of costs in individual years after the diagnosis, and to provide reference costs to which discount rates can be applied, I charted excess costs with confidence intervals and also presented tables of mean excess costs. Excess costs in individual years were calculated as arithmetic means of the differences in total costs between matched pairs. This method was inappropriate for testing the equality of care phases, because paired individuals did not share the same phase trajectories. Consequently, univariable GLM regressions were performed to test the equality of costs across phases and provide reference costs with confidence intervals. The model specifications were as above, but were performed on total costs over a particular phase rather than the entire eight-year follow-up. I explored the possibility of using two-part models, which use conditional regression incorporating both binary and continuous outcome estimations to cope with high zero counts. However, for binary outcomes with matched samples, standard binary outcome regression models are inappropriate, and how the two-part models could be specified to produce robust estimates on matched data was outside my statistical knowledge. Therefore, I again used GLMs with gamma distribution and log

link, in line with the other models, and estimated excess costs from average marginal effects.

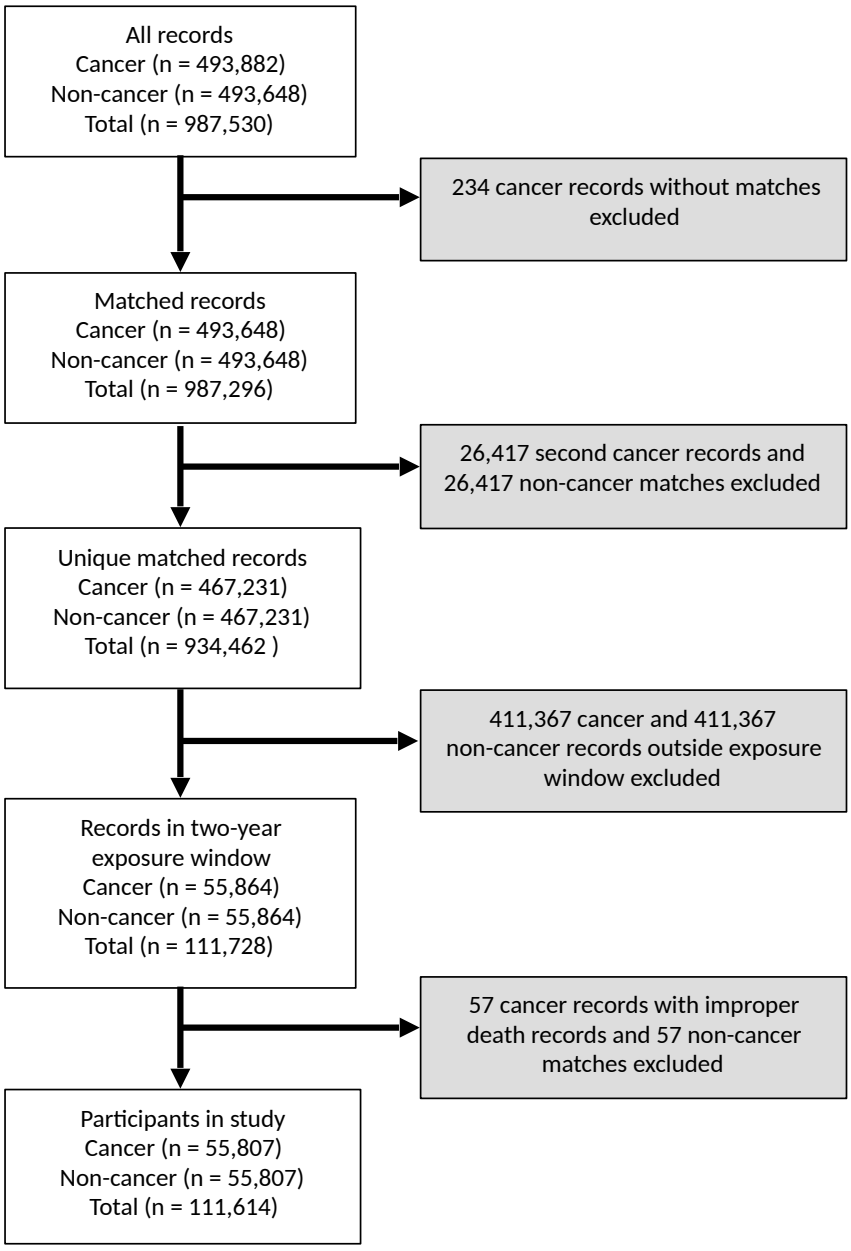
All charts and tables were stratified by the 10 most common cancer types listed by NHS Scotland [24] and all other cancers combined. In all statistical tests 5% significance levels were used and I reported results with 95% confidence intervals where appropriate. Cost calculations were carried out using Stata's internal float precision, but final costs were reported rounded to the nearest pound sterling at 2017/18 price levels. All analyses were carried out in Stata 16.0 and cost figures were rounded in Microsoft Excel.

4.3 Results

4.3.1 Participants

From a total of 493,882 cancer records in SMR06, 493,648 records of individuals without any history of cancer were matched. After the removal of 234 unmatched cancer cases, second cancers, individuals under 18 years of age, and records outside the entry period, 55,807 cancer patients and 55,807 individuals without any history of cancer were included in the final analysis. A flow chart showing more detail is given in Figure 4.1.

Figure 4.1.: Flow chart of participant numbers



Notes: Improper death certificates were only appropriate to SMR06 records so did not occur in the non-cancer group. Second cancer records were also only relevant to cancer patients. Non-cancer records were given a pseudo-diagnosis date equal to their cancer matches, hence record numbers outside the exposure window were necessarily identical for both cohorts.

4.3.2 Descriptive Data

Participant characteristics at the time of the diagnosis-event are shown in Table 4.2 and Table 4.3, and indicate precise matching for age, sex, SIMD quintile and NHS Health Board. The mean age for both cohorts was 67.5 with 37.8% under 65 years of age, while 51.7% of all individuals were female. The West Regional Network, which

includes the Glasgow conurbation, contained the highest proportion of individuals with 47.2% compared to 27.6% for South and East, and 25.2% for North. Comorbidities varied between cancer sites and across cohorts. Across all cancer sites combined, 9.3% of the non-cancer group had one comorbidity compared to 13.4% of cancer patients, while 3.1% of the non-cancer group had two or more comorbidities compared to 4.2% of cancer.

Table 4.2.: Characteristics of the cancer and non-cancer cohorts at baseline part 1

	Trachea, bronchus & lung		Breast		Colorectal		Prostate		Head & neck		Malignant melanoma of skin	
Variable	No cancer N=9,132	Cancer N=9,132	No cancer N=8,138	Cancer N=8,138	No cancer N=7,270	Cancer N=7,270	No cancer N=5,770	Cancer N=5,770	No cancer N=2,130	Cancer N=2,130	No cancer N=2,062	Cancer N=2,062
Sex												
Male	4,718 (51.7%)	4,718 (51.7%)	46 (0.6%)	46 (0.6%)	3,942 (54.2%)	3,942 (54.2%)	5,770 (100.0%)	5,770 (100.0%)	1,481 (69.5%)	1,481 (69.5%)	905 (43.9%)	905 (43.9%)
Female	4,414 (48.3%)	4,414 (48.3%)	8,092 (99.4%)	8,092 (99.4%)	3,328 (45.8%)	3,328 (45.8%)	NA	NA	649 (30.5%)	649 (30.5%)	1,157 (56.1%)	1,157 (56.1%)
Age in years (mean, sd)	71.4 (10.6)	71.4 (10.6)	62.6 (14.1)	62.6 (14.1)	70.4 (11.8)	70.4 (11.8)	70.6 (9.4)	70.6 (9.4)	63.7 (12.3)	63.7 (12.3)	59.5 (17.5)	59.5 (17.5)
Age categories												
< 50	249 (2.7%)	249 (2.7%)	1,551 (19.1%)	1,551 (19.1%)	325 (4.5%)	325 (4.5%)	55 (1.0%)	55 (1.0%)	249 (11.7%)	249 (11.7%)	635 (30.8%)	635 (30.8%)
50-59	996 (10.9%)	996 (10.9%)	1,896 (23.3%)	1,896 (23.3%)	952 (13.1%)	952 (13.1%)	599 (10.4%)	599 (10.4%)	500 (23.5%)	500 (23.5%)	355 (17.2%)	355 (17.2%)
60-69	2,396 (26.2%)	2,396 (26.2%)	2,191 (26.9%)	2,191 (26.9%)	1,909 (26.3%)	1,909 (26.3%)	2,030 (35.2%)	2,030 (35.2%)	683 (32.1%)	683 (32.1%)	396 (19.2%)	396 (19.2%)
70-79	3,334 (36.5%)	3,334 (36.5%)	1,396 (17.2%)	1,396 (17.2%)	2,379 (32.7%)	2,379 (32.7%)	2,026 (35.1%)	2,026 (35.1%)	479 (22.5%)	479 (22.5%)	400 (19.4%)	400 (19.4%)
>= 80	2,157 (23.6%)	2,157 (23.6%)	1,100 (13.5%)	1,100 (13.5%)	1,701 (23.4%)	1,701 (23.4%)	1,059 (18.4%)	1,059 (18.4%)	219 (10.3%)	219 (10.3%)	276 (13.4%)	276 (13.4%)
Age under 65	2,247 (24.6%)	2,247 (24.6%)	4,619 (56.8%)	4,619 (56.8%)	2,162 (29.7%)	2,162 (29.7%)	1,588 (27.5%)	1,588 (27.5%)	1,121 (52.6%)	1,121 (52.6%)	1,202 (58.3%)	1,202 (58.3%)
SIMD quintile												
Least deprived	1,005 (11.0%)	1,005 (11.0%)	1,757 (21.6%)	1,757 (21.6%)	1,389 (19.1%)	1,389 (19.1%)	1,306 (22.6%)	1,306 (22.6%)	247 (11.6%)	247 (11.6%)	559 (27.1%)	559 (27.1%)
2nd least deprived	1,339 (14.7%)	1,339 (14.7%)	1,755 (21.6%)	1,755 (21.6%)	1,437 (19.8%)	1,437 (19.8%)	1,273 (22.1%)	1,273 (22.1%)	350 (16.4%)	350 (16.4%)	448 (21.7%)	448 (21.7%)
3rd least deprived	1,734 (19.0%)	1,734 (19.0%)	1,660 (20.4%)	1,660 (20.4%)	1,522 (20.9%)	1,522 (20.9%)	1,238 (21.5%)	1,238 (21.5%)	409 (19.2%)	409 (19.2%)	435 (21.1%)	435 (21.1%)
2nd most deprived	2,256 (24.7%)	2,256 (24.7%)	1,568 (19.3%)	1,568 (19.3%)	1,553 (21.4%)	1,553 (21.4%)	1,089 (18.9%)	1,089 (18.9%)	488 (22.9%)	488 (22.9%)	338 (16.4%)	338 (16.4%)
Most deprived	2,798 (30.6%)	2,798 (30.6%)	1,398 (17.2%)	1,398 (17.2%)	1,369 (18.8%)	1,369 (18.8%)	864 (15.0%)	864 (15.0%)	636 (29.9%)	636 (29.9%)	282 (13.7%)	282 (13.7%)
Regional network												
South and east	2,463 (27.0%)	2,463 (27.0%)	2,168 (26.6%)	2,168 (26.6%)	2,032 (28.0%)	2,032 (28.0%)	1,752 (30.4%)	1,752 (30.4%)	564 (26.5%)	564 (26.5%)	540 (26.2%)	540 (26.2%)
West	4,684 (51.3%)	4,684 (51.3%)	3,820 (46.9%)	3,820 (46.9%)	3,284 (45.2%)	3,284 (45.2%)	2,498 (43.3%)	2,498 (43.3%)	1,089 (51.1%)	1,089 (51.1%)	997 (48.4%)	997 (48.4%)
North	1,985 (21.7%)	1,985 (21.7%)	2,150 (26.4%)	2,150 (26.4%)	1,954 (26.9%)	1,954 (26.9%)	1,520 (26.3%)	1,520 (26.3%)	477 (22.4%)	477 (22.4%)	525 (25.5%)	525 (25.5%)
Comorbidity count												
Zero	7,793 (85.3%)	6,440 (70.5%)	7,412 (91.1%)	7,415 (91.1%)	6,301 (86.7%)	6,167 (84.8%)	4,969 (86.1%)	5,058 (87.7%)	1,908 (89.6%)	1,796 (84.3%)	1,881 (91.2%)	1,901 (92.2%)
One	1,009 (11.0%)	1,996 (21.9%)	556 (6.8%)	577 (7.1%)	744 (10.2%)	853 (11.7%)	569 (9.9%)	547 (9.5%)	173 (8.1%)	253 (11.9%)	138 (6.7%)	125 (6.1%)
Two or more	330 (3.6%)	696 (7.6%)	170 (2.1%)	146 (1.8%)	225 (3.1%)	250 (3.4%)	232 (4.0%)	165 (2.9%)	49 (2.3%)	81 (3.8%)	43 (2.1%)	36 (1.7%)
Acute myocardial infarcation	221 (2.4%)	266 (2.9%)	78 (1.0%)	80 (1.0%)	158 (2.2%)	135 (1.9%)	150 (2.6%)	124 (2.1%)	38 (1.8%)	50 (2.3%)	32 (1.6%)	25 (1.2%)
Congestive heart failure	135 (1.5%)	192 (2.1%)	62 (0.8%)	67 (0.8%)	100 (1.4%)	104 (1.4%)	94 (1.6%)	55 (1.0%)	19 (0.9%)	14 (0.7%)	12 (0.6%)	19 (0.9%)
Peripheral vascular disease	130 (1.4%)	292 (3.2%)	47 (0.6%)	49 (0.6%)	74 (1.0%)	103 (1.4%)	76 (1.3%)	83 (1.4%)	21 (1.0%)	<30	10 (0.5%)	<30
Cerebral vascular disease	198 (2.2%)	297 (3.3%)	105 (1.3%)	82 (1.0%)	139 (1.9%)	104 (1.4%)	130 (2.3%)	118 (2.0%)	32 (1.5%)	56 (2.6%)	33 (1.6%)	18 (0.9%)
Dementia	142 (1.6%)	168 (1.8%)	92 (1.1%)	83 (1.0%)	95 (1.3%)	97 (1.3%)	60 (1.0%)	54 (0.9%)	18 (0.8%)	16 (0.8%)	18 (0.9%)	13 (0.6%)
Chronic pulmonary disease	349 (3.8%)	1,404 (15.4%)	254 (3.1%)	226 (2.8%)	257 (3.5%)	279 (3.8%)	194 (3.4%)	172 (3.0%)	69 (3.2%)	116 (5.4%)	56 (2.7%)	48 (2.3%)
Rheumatoid disease	65 (0.7%)	101 (1.1%)	52 (0.6%)	35 (0.4%)	46 (0.6%)	38 (0.5%)	14 (0.2%)	14 (0.2%)	<10	<10	<10	<10
Peptic ulcer	50 (0.5%)	40 (0.4%)	15 (0.2%)	24 (0.3%)	31 (0.4%)	33 (0.5%)	30 (0.5%)	11 (0.2%)	<10	12 (0.6%)	<10	<10
Diabetes	284 (3.1%)	431 (4.7%)	147 (1.8%)	151 (1.9%)	216 (3.0%)	314 (4.3%)	208 (3.6%)	154 (2.7%)	43 (2.0%)	69 (3.2%)	37 (1.8%)	39 (1.9%)
Renal disease - moderate or severe	109 (1.2%)	209 (2.3%)	47 (0.6%)	63 (0.8%)	83 (1.1%)	112 (1.5%)	67 (1.2%)	89 (1.5%)	12 (0.6%)	12 (0.6%)	13 (0.6%)	10 (0.5%)

Notes: SIMD = Scottish Index of Multiple Deprivation, sd = standard deviation
. Stratified by the 10 most common cancers recorded by NHS Scotland.
Comorbidity measures were recorded in SMR01 prior to SMR06 registration.
Individual health boards were aggregated to network regions due to many low counts for remote regions.
Entries with low counts are rounded to prevent disclosure from differencing across rows or columns.

Table 4.3.: Characteristics of the cancer and non-cancer cohorts at baseline part 2

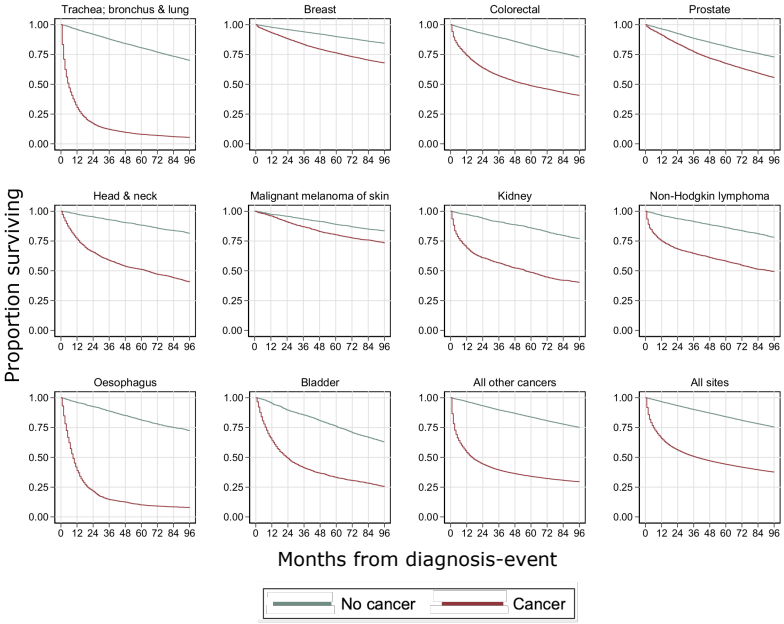
	Kidney		Non-Hodgkin lymphoma		Oesophagus		Bladder		All other cancers		All cancers	
	No cancer N=1,541	Cancer N=1,541	No cancer N=1,881	Cancer N=1,881	No cancer N=1,590	Cancer N=1,590	No cancer N=1,373	Cancer N=1,373	No cancer N=14,920	Cancer N=14,920	No cancer N=5,5807	Cancer N=55,807
Sex												
Male	877 (56.9%)	877 (56.9%)	932 (49.5%)	932 (49.5%)	1,004 (63.1%)	1,004 (63.1%)	910 (66.3%)	910 (66.3%)	6,350 (42.6%)	6,350 (42.6%)	26,935 (48.3%)	26,935 (48.3%)
Female	664 (43.1%)	664 (43.1%)	949 (50.5%)	949 (50.5%)	586 (36.9%)	586 (36.9%)	463 (33.7%)	463 (33.7%)	8,570 (57.4%)	8,570 (57.4%)	28,872 (51.7%)	28,872 (51.7%)
Age in years (mean, sd)	67.6 (12.9)	67.6 (12.9)	66.3 (14.2)	66.3 (14.2)	70.2 (11.6)	70.2 (11.6)	73.8 (10.9)	73.8 (10.9)	66.2 (16.1)	66.2 (16.1)	67.5 (13.8)	67.5 (13.8)
Age categories												
< 50	140 (9.1%)	140 (9.1%)	229 (12.2%)	229 (12.2%)	67 (4.2%)	67 (4.2%)	35 (2.5%)	35 (2.5%)	2,348 (15.7%)	2,348 (15.7%)	5,883 (10.5%)	5,883 (10.5%)
50-59	261 (16.9%)	261 (16.9%)	316 (16.8%)	316 (16.8%)	226 (14.2%)	226 (14.2%)	103 (7.5%)	103 (7.5%)	2,015 (13.5%)	2,015 (13.5%)	8,219 (14.7%)	8,219 (14.7%)
60-69	417 (27.1%)	417 (27.1%)	487 (25.9%)	487 (25.9%)	446 (28.1%)	446 (28.1%)	303 (22.1%)	303 (22.1%)	3,351 (22.5%)	3,351 (22.5%)	14,609 (26.2%)	14,609 (26.2%)
70-79	428 (27.8%)	428 (27.8%)	504 (26.8%)	504 (26.8%)	481 (30.3%)	481 (30.3%)	463 (33.7%)	463 (33.7%)	3,969 (26.6%)	3,969 (26.6%)	15,859 (28.4%)	15,859 (28.4%)
>= 80	295 (19.1%)	295 (19.1%)	345 (18.3%)	345 (18.3%)	370 (23.3%)	370 (23.3%)	469 (34.2%)	469 (34.2%)	3,237 (21.7%)	3,237 (21.7%)	11,237 (20.1%)	11,237 (20.1%)
Age under 65	628 (40.8%)	628 (40.8%)	805 (42.8%)	805 (42.8%)	513 (32.3%)	513 (32.3%)	263 (19.2%)	263 (19.2%)	5,951 (39.9%)	5,951 (39.9%)	21,099 (37.8%)	21,099 (37.8%)
SIMD quintile												
Least deprived	274 (17.8%)	274 (17.8%)	392 (20.8%)	392 (20.8%)	234 (14.7%)	234 (14.7%)	238 (17.3%)	238 (17.3%)	2,627 (17.6%)	2,627 (17.6%)	10,028 (18.0%)	10,028 (18.0%)
2nd least deprived	274 (17.8%)	274 (17.8%)	368 (19.6%)	368 (19.6%)	289 (18.2%)	289 (18.2%)	248 (18.1%)	248 (18.1%)	2,871 (19.2%)	2,871 (19.2%)	10,652 (19.1%)	10,652 (19.1%)
3rd least deprived	331 (21.5%)	331 (21.5%)	404 (21.5%)	404 (21.5%)	337 (21.2%)	337 (21.2%)	303 (22.1%)	303 (22.1%)	3,061 (20.5%)	3,061 (20.5%)	11,434 (20.5%)	11,434 (20.5%)
2nd most deprived	343 (22.3%)	343 (22.3%)	377 (20.0%)	377 (20.0%)	355 (22.3%)	355 (22.3%)	288 (21.0%)	288 (21.0%)	3,228 (21.6%)	3,228 (21.6%)	11,883 (21.3%)	11,883 (21.3%)
Most deprived	319 (20.7%)	319 (20.7%)	340 (18.1%)	340 (18.1%)	375 (23.6%)	375 (23.6%)	296 (21.6%)	296 (21.6%)	3,133 (21.0%)	3,133 (21.0%)	11,810 (21.2%)	11,810 (21.2%)
Regional network												
South and east	433 (28.1%)	433 (28.1%)	559 (29.7%)	559 (29.7%)	403 (25.3%)	403 (25.3%)	387 (28.2%)	387 (28.2%)	4,101 (27.5%)	4,101 (27.5%)	15,402 (27.6%)	15,402 (27.6%)
West	715 (46.4%)	715 (46.4%)	818 (43.5%)	818 (43.5%)	761 (47.9%)	761 (47.9%)	603 (43.9%)	603 (43.9%)	7,099 (47.6%)	7,099 (47.6%)	26,368 (47.2%)	26,368 (47.2%)
North	393 (25.5%)	393 (25.5%)	504 (26.8%)	504 (26.8%)	426 (26.8%)	426 (26.8%)	383 (27.9%)	383 (27.9%)	3,720 (24.9%)	3,720 (24.9%)	14,037 (25.2%)	14,037 (25.2%)
Comorbidity count												
Zero	1,375 (89.2%)	1,243 (80.7%)	1,676 (89.1%)	1,565 (83.2%)	1,363 (85.7%)	1,295 (81.4%)	1,153 (84.0%)	1,101 (80.2%)	13,093 (87.8%)	12,000 (80.4%)	48,924 (87.7%)	45,981 (82.4%)
One	134 (8.7%)	218 (14.1%)	151 (8.0%)	239 (12.7%)	173 (10.9%)	237 (14.9%)	164 (11.9%)	201 (14.6%)	1,368 (9.2%)	2,247 (15.1%)	5,179 (9.3%)	7,493 (13.4%)
Two or more	32 (2.1%)	80 (5.2%)	54 (2.9%)	77 (4.1%)	54 (3.4%)	58 (3.6%)	56 (4.1%)	71 (5.2%)	459 (3.1%)	673 (4.5%)	1,704 (3.1%)	2,333 (4.2%)
Acute myocardial infarcation	29 (1.9%)	43 (2.8%)	43 (2.3%)	34 (1.8%)	41 (2.6%)	35 (2.2%)	39 (2.8%)	37 (2.7%)	292 (2.0%)	287 (1.9%)	1,121 (2.0%)	1,116 (2.0%)
Congestive heart failure	19 (1.2%)	42 (2.7%)	23 (1.2%)	26 (1.4%)	30 (1.9%)	22 (1.4%)	28 (2.0%)	22 (1.6%)	187 (1.3%)	230 (1.5%)	709 (1.3%)	793 (1.4%)
Peripheral vascular disease	14 (0.9%)	31 (2.0%)	15 (0.8%)	<30	17 (1.1%)	<30	17 (1.2%)	<30	142 (1.0%)	186 (1.2%)	563 (1.0%)	867 (1.6%)
Cerebral vascular disease	22 (1.4%)	24 (1.6%)	29 (1.5%)	29 (1.5%)	33 (2.1%)	18 (1.1%)	42 (3.1%)	22 (1.6%)	315 (2.1%)	311 (2.1%)	1,078 (1.9%)	1,079 (1.9%)
Dementia	15 (1.0%)	23 (1.5%)	20 (1.1%)	18 (1.0%)	24 (1.5%)	22 (1.4%)	22 (1.6%)	23 (1.7%)	200 (1.3%)	250 (1.7%)	706 (1.3%)	767 (1.4%)
Chronic pulmonary disease	37 (2.4%)	60 (3.9%)	55 (2.9%)	74 (3.9%)	51 (3.2%)	92 (5.8%)	51 (3.7%)	79 (5.8%)	486 (3.3%)	676 (4.5%)	1,859 (3.3%)	3,226 (5.8%)
Rheumatoid disease	<10	10 (0.6%)	12 (0.6%)	19 (1.0%)	16 (1.0%)	<10	<10	<10	84 (0.6%)	103 (0.7%)	317 (0.6%)	349 (0.6%)
Peptic ulcer	<10	<10	<10	20 (1.1%)	<10	19 (1.2%)	<10	<10	48 (0.3%)	159 (1.1%)	200 (0.4%)	334 (0.6%)
Diabetes	36 (2.3%)	69 (4.5%)	41 (2.2%)	75 (4.0%)	47 (3.0%)	61 (3.8%)	46 (3.4%)	72 (5.2%)	388 (2.6%)	735 (4.9%)	1,493 (2.7%)	2,170 (3.9%)
Renal disease - moderate or severe	11 (0.7%)	56 (3.6%)	12 (0.6%)	45 (2.4%)	19 (1.2%)	23 (1.4%)	27 (2.0%)	55 (4.0%)	154 (1.0%)	317 (2.1%)	554 (1.0%)	991 (1.8%)

Notes: SIMD = Scottish Index of Multiple Deprivation, sd = standard deviation.
Stratified by the 10 most common cancers recorded by NHS Scotland.
Comorbidity measures were recorded in SMR01 prior to SMR06 registration.
Individual health boards were aggregated to network regions due to many low counts for remote regions.
Entries with low counts are rounded to prevent disclosure from differencing across rows or columns.

4.3.3 Trajectories of Outcomes

Figure 4.2 compares survival in the cohorts over time. As expected, survival was poorer for the cancer cohort in all types of cancer, with around one-third of patients surviving the eight-year period compared to approximately three-quarters of individuals without cancer. Furthermore, the cancer cohort showed considerable variation across cancer sites, with approximately three-quarters of patients with skin melanoma alive at the end of follow-up compared to approximately 10% for oesophagus cancer and lung-related cancers, while the non-cancer cohort showed notably lower variation across the matched cancer types.

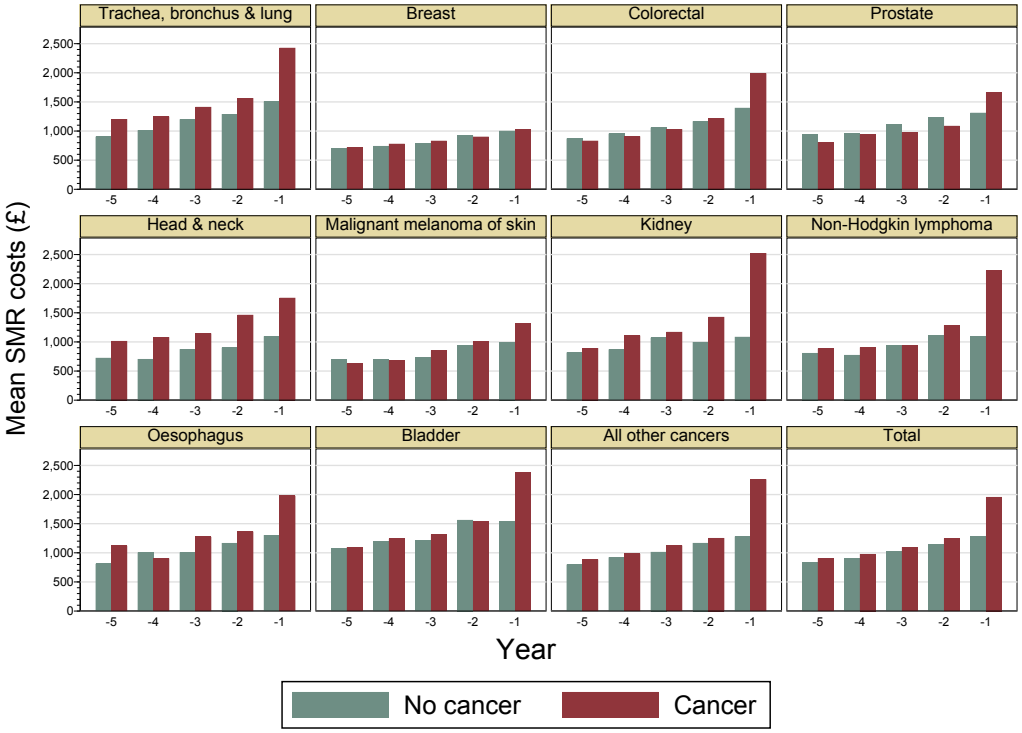
Figure 4.2.: Kaplan-Meier trajectories of survival for the cancer and non-cancer cohorts



Note: Dates of deaths are rounded to the nearest month.

Figure 4.3 shows the sum of inpatient and outpatient costs in the pre-diagnosis period, with costs in both groups rising throughout the pre-diagnosis. Care must be taken when comparing pre-diagnosis trends with post-diagnosis trends where deaths over time would affect costs. This could not occur in the pre-diagnosis period as all individuals were alive at the baseline diagnosis event. Prior to the year before diagnosis, the cancer and non-cancer groups had similar costs for most cancers, although lung-related cancers and head and neck cancers appear elevated at all years. During the year before the diagnosis-event, costs appear notably elevated in the cancer cohort, suggesting excess costs may not be confined to the post-diagnosis period.

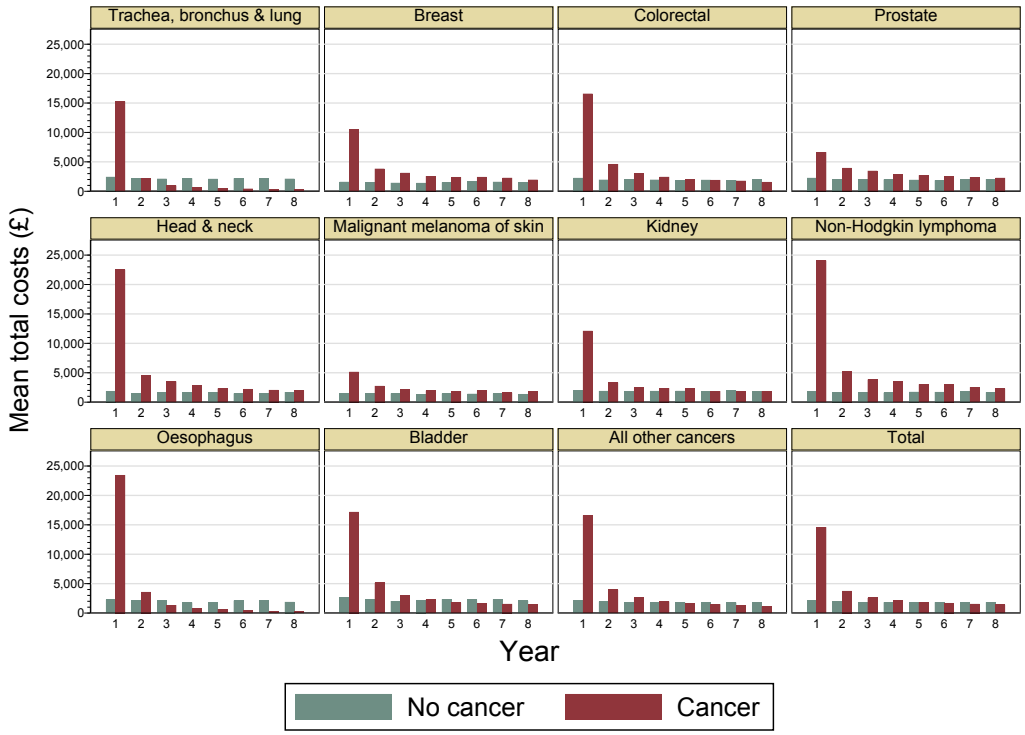
Figure 4.3.: Trajectories of costs by year by and cancer type in the pre-diagnosis period for the cancer and non-cancer cohorts



Notes: SMR = Scottish Morbidity Record. Pre-diagnosis costs comprised SMR00/01 costs only and excluded prescriptions. Years are relative to the diagnosis-event. All costs are undiscounted at 2018 price levels.

Yearly post-diagnosis trajectories in Figure 4.4, show clear differences in shape between the cohorts, with that of the non-cancer group largely flat compared to a large spike in year 1 and exponential declines thereafter in the cancer group. Higher variation across cancer sites can be seen in the cancer group than the non-cancer group, although the smaller magnitude of costs in the latter may make differences less apparent. Costs for breast cancer, prostate cancer, head and neck cancer, skin cancer, and non-Hodgkin lymphoma stayed elevated above their controls throughout the eight-year follow-up, while trachea, bronchus and lung cancers, oesophagus cancers, bladder cancers and all other cancers dropped below their controls.

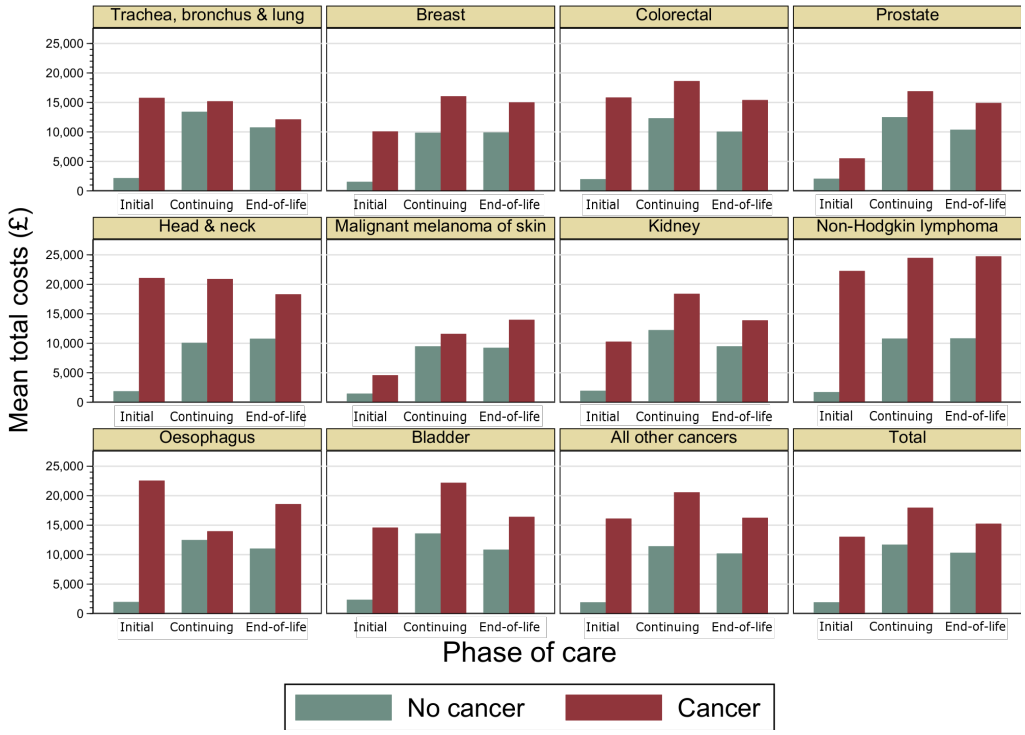
Figure 4.4.: Trajectories of total costs by year and by cancer type in the post-diagnosis period for the cancer and non-cancer cohorts



Notes: Costs are the sum of inpatient/daycase, outpatient and prescriptions. Years are relative to the diagnosis-event. All costs are undiscounted at 2018 price levels.

Phase-of-care trajectories in Figure 4.5 show apparently higher costs in the cancer group in all phases across all cancer sites, with the most notable difference between cohorts occurring during the initial phase. Notable variation between cancer types was observed in cancer patients, both in the shape and magnitude of costs. The non-cancer group showed less variation across matched cancer sites, with low costs in the initial period and substantially higher costs in the continuing and end-of-life periods.

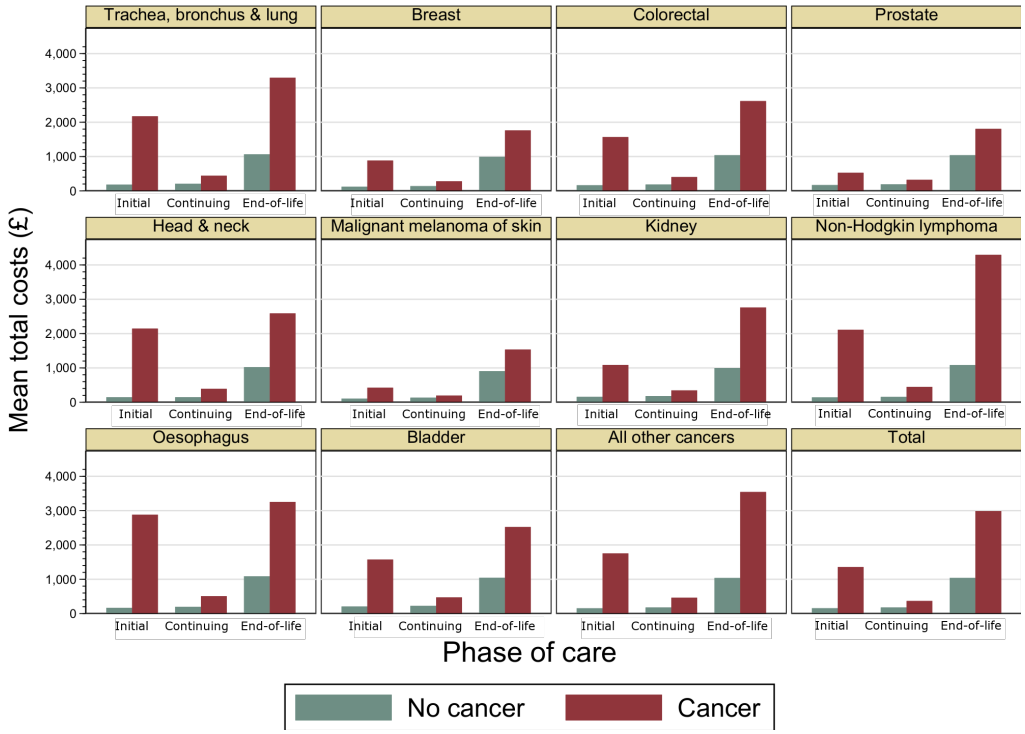
Figure 4.5.: Trajectories of total costs by phase-of-care and by cancer type for the cancer and non-cancer cohorts



Notes: Phases are relative to the diagnosis-event. Total costs are the sum of inpatient/daycase, outpatient and prescriptions. All costs are undiscounted at 2018 price levels.

Variation across phases may have been driven to some extent by the differing duration of phases, with the continuing period potentially being of considerably greater length. An alternative view presented in Figure 4.6 shows the monthly rate of cost accrual in the post-diagnosis phases. The highest rates of cost accrual occurred in the end-of-life period for both cohorts across all cancer sites, with the initial period next for the cancer cohort and the continuing period notably lower. In the non-cancer cohort rates appeared similar across the initial and continuous periods, and the magnitudes were low relative to the cancer cohort, which may obscure variation across phases. Once again, the cancer cohort had higher costs in all phases across all cancers, and little variation was observed across cancer sites in the non-cancer cohort.

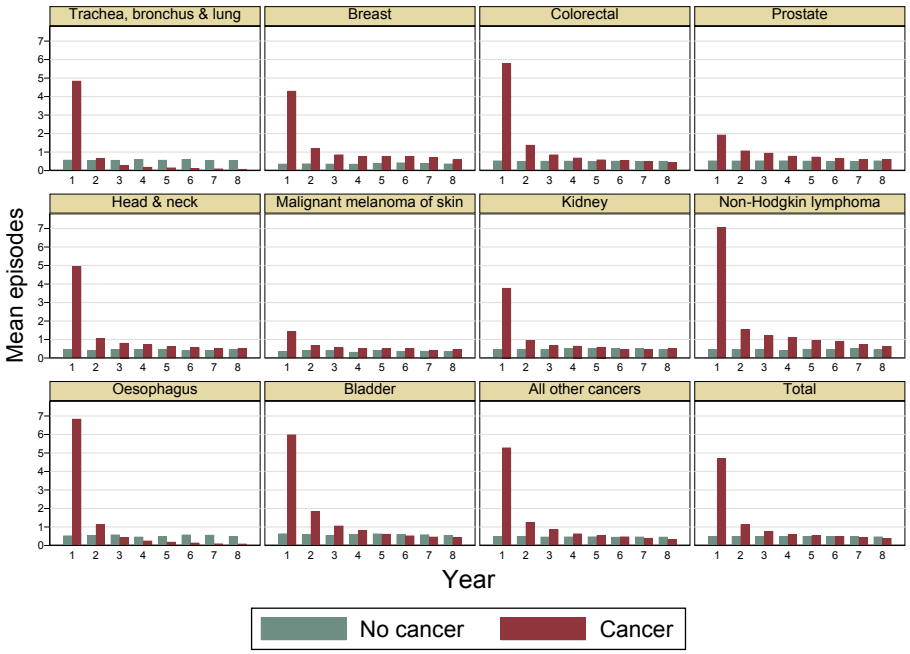
Figure 4.6.: Trajectories of monthly total costs by phase-of-care and by cancer type for the cancer and non-cancer cohorts



Notes: Monthly costs = costs in phase / months spent in phase. Costs are the sum of inpatient/daycase, outpatient and prescriptions. Phases are relative to the diagnosis-event. All costs are undiscounted at 2018 price levels.

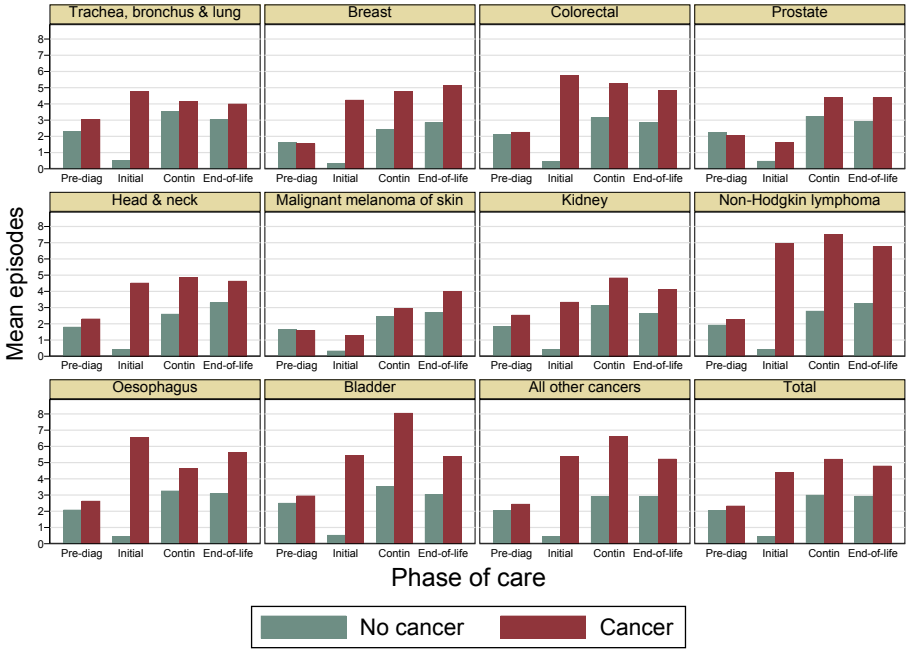
Figures 4.7–4.10 show inpatient episodes and days, stratified by year after diagnosis and phase-of-care. The phase-of-care charts represent only living patients, consequently, the pre-diagnosis period was included. In this phase, both cohorts showed similar levels of resource use within cancer sites, although for some sites the cancer cohort was slightly elevated. The initial phase contained notably higher episodes and days in the cancer cohort, while the continuing and end-of-life phases had notably higher episodes without correspondingly higher numbers of days, suggesting that cancer patients accrued more frequent but shorter inpatient stays. The annual trends were similar to those of annual costs, likely reflecting the sizeable contribution of inpatient episodes to the total costs described in Section 3.3.

Figure 4.7.: Trajectories of inpatient episodes by year and by cancer type for the cancer and non-cancer cohorts



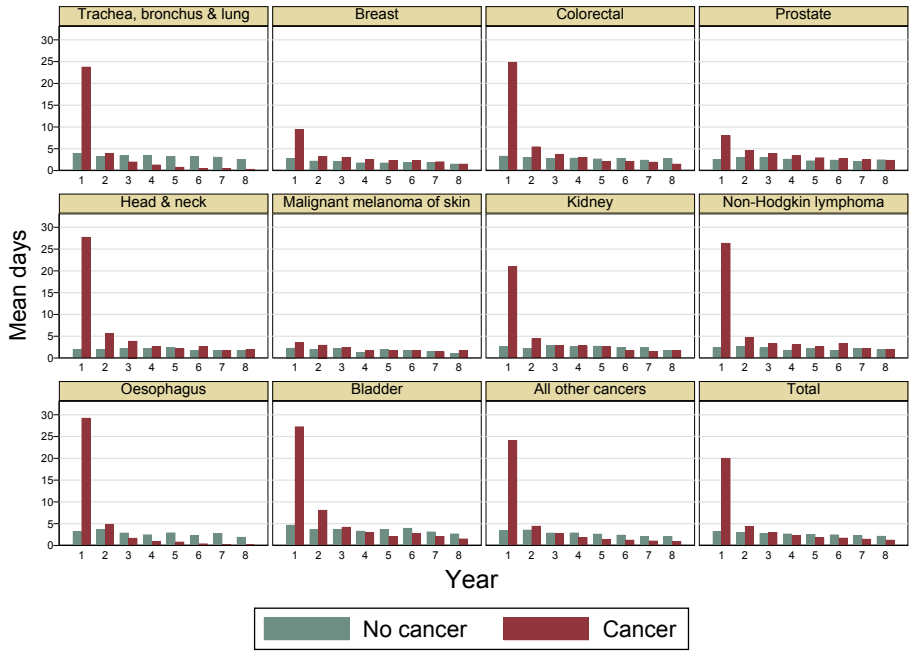
Notes: Years are relative to the diagnosis-event. Inpatient episodes included daycase episodes.

Figure 4.8.: Trajectories of inpatient episodes by phase-of-care and by cancer type for the cancer and non-cancer cohorts



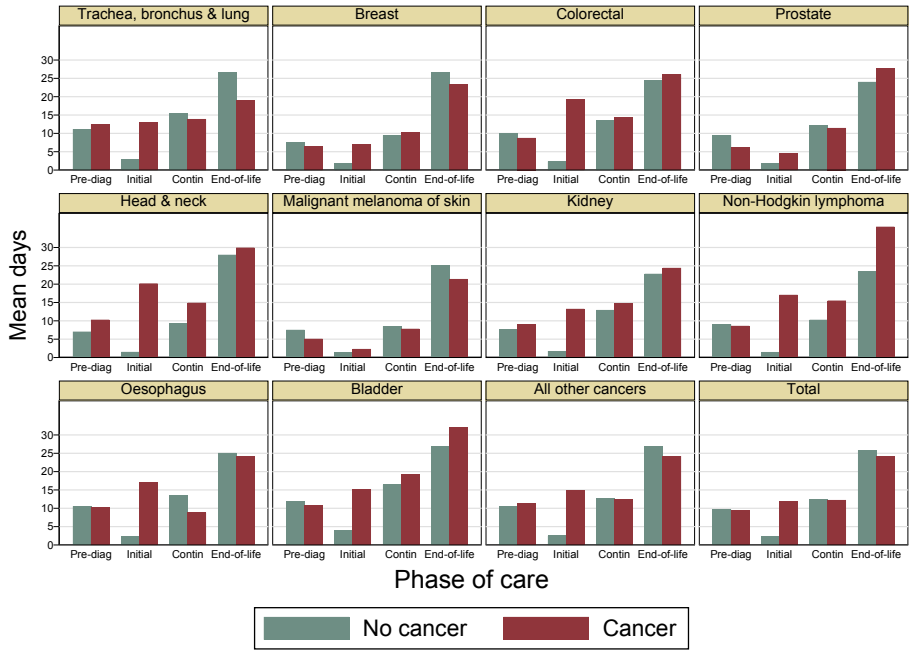
Notes: Pre-diag = pre-diagnosis, Contin = continuing. Phases are relative to the diagnosis-event. Inpatient episodes included daycase episodes.

Figure 4.9.: Trajectories of inpatient days by year and by cancer type for the cancer and non-cancer cohorts



Notes: Included daycase episodes where 1 daycase episode = 1 day. Years are relative to the diagnosis-event. For longer episodes days were calculated per-episode rather than per-stay.

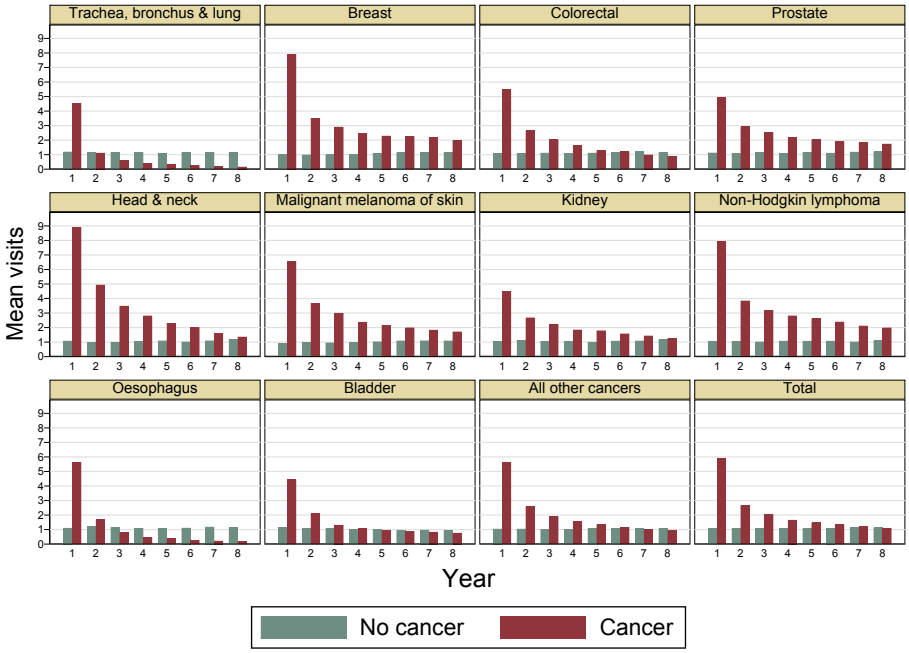
Figure 4.10.: Trajectories of inpatient days by phase-of-care and by cancer type for the cancer and non-cancer cohorts



Notes: Pre-diag = pre-diagnosis, Contin = continuing. Included daycase episodes where 1 daycase episode = 1 day. For longer episodes days were calculated per-episode rather than per-stay.

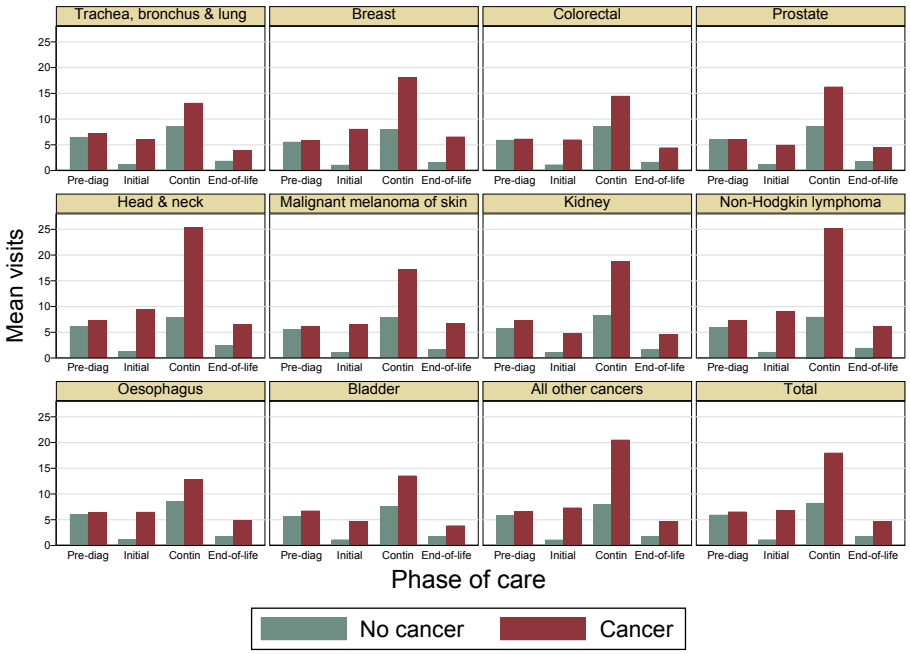
Outpatient visits shown in figures 4.11–4.12 took a similar pattern to inpatient episodes, however the initial and end-of-life phases had relatively lower contributions in both cohorts, with peak visits seen during the continuing period.

Figure 4.11.: Trajectories of outpatient visits by year and by cancer type for the cancer and non-cancer cohorts



Notes: Years are relative to the diagnosis-event.

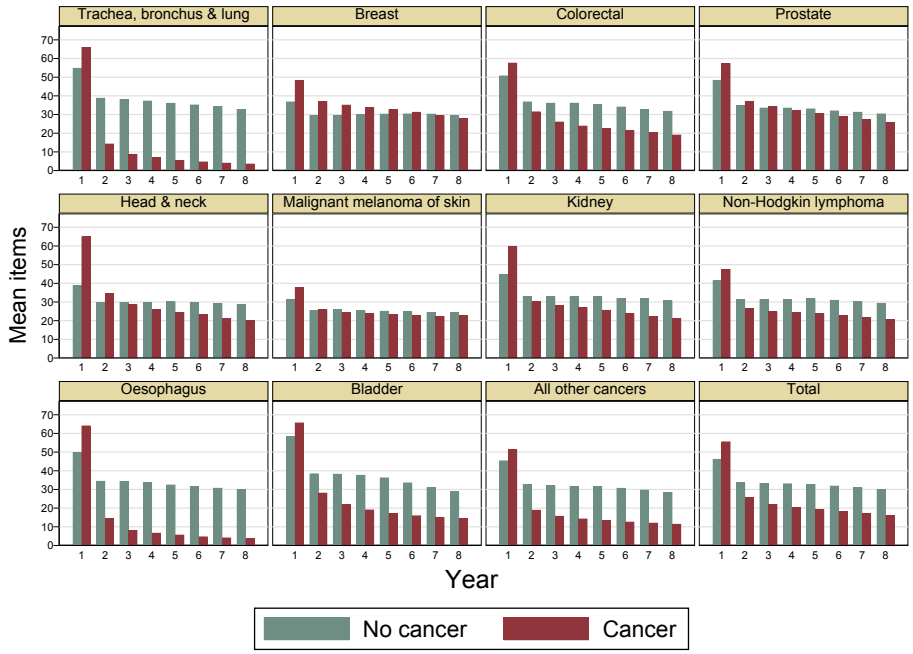
Figure 4.12.: Trajectories of outpatient visits by phase-of-care and by cancer type for the cancer and non-cancer cohorts



Notes: Pre-diag = pre-diagnosis, Contin = continuing. Phases are relative to the diagnosis-event.

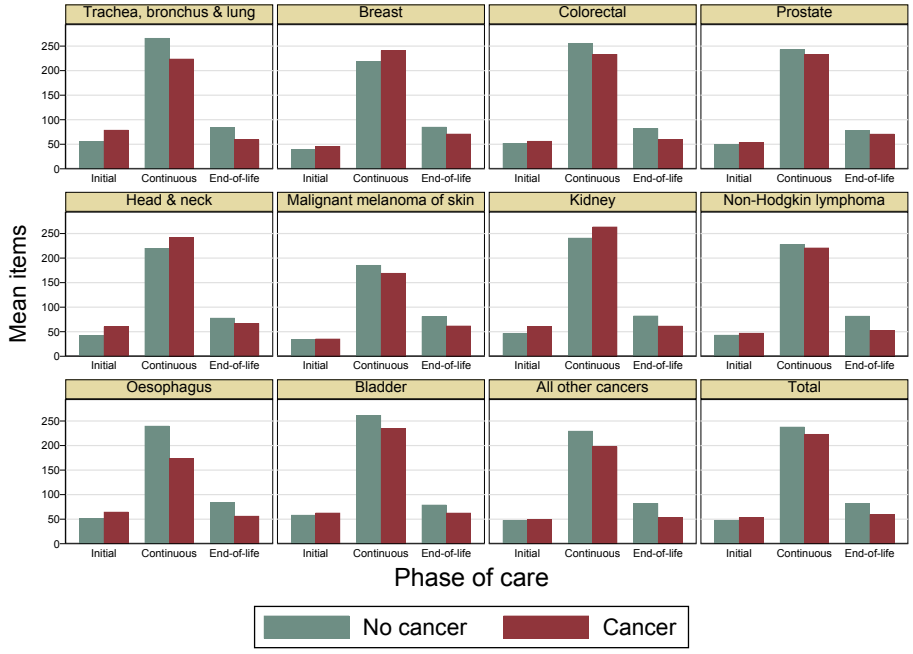
Numbers of prescribed items shown in figures 4.13–4.14 followed notably different trajectories from those of hospital visits, with much smaller peaks for the cancer cohort in the first year after diagnosis. No strong differences were observed between the cancer and non-cancer cohorts in any phase, except perhaps the end-of-life phase with higher levels seen in the non-cancer cohort, which may have reflected longer survival in this phase. The continuing phase contained notably higher numbers of items in both cohorts.

Figure 4.13.: Trajectories of prescribed items by year and by cancer type for the cancer and non-cancer cohorts



Note: All numbers refer to individual prescribed items. Years are relative to the diagnosis-event.

Figure 4.14.: Trajectories of prescribed items by phase-of-care and by cancer type for the cancer and non-cancer cohorts



Note: Pre-diag = pre-diagnosis, Contin = continuing. All numbers refer to individual prescribed items. Phases are relative to the diagnosis-event.

4.3.4 GLM Estimations of Excess Costs

Table 4.4 shows that the eight-year excess costs were positive and substantial for all cancer types analysed in both univariable and multivariable estimates. Adjustment for unmatched confounders raised the magnitudes of excess costs for all cancers but not substantially. The highest excess costs, both in absolute monetary units and as a ratio of the comparison group's costs, were observed for non-Hodgkin lymphoma (adjusted mean £34,106; 95%CI £31,623 to £36,589) with the lowest found for trachea, bronchus and lung cancers (adjusted mean £3,153; 95%CI £2,484 to £3,821). Full outputs of the regression models, including additional confounders to those matched for, are shown in Table 4.5.

Table 4.4.: Univariable and multivariable GLM estimates of eight-year post-diagnosis excess costs of cancer

	Univariable				Multivariable			
	Value	*p*	95% CI		Value	*p*	95% CI	
All cancers								
Cost ratio	1.960	<0.001	1.930	1.989	1.985	<0.001	1.956	2.015
Excess cost (£)	14,427	<0.001	14,105	14,749	14,717	<0.001	14,393	15,042
Trachea, bronchus & lung								
Cost ratio	1.178	<0.001	1.137	1.221	1.180	<0.001	1.138	1.223
Excess cost (£)	3,133	<0.001	2,462	3,804	3,153	<0.001	2,484	3,821
Breast								
Cost ratio	2.392	<0.001	2.297	2.491	2.495	<0.001	2.396	2.597
Excess cost (£)	16,942	<0.001	16,202	17,682	17,817	<0.001	17,049	18,585
Colorectal								
Cost ratio	2.135	<0.001	2.057	2.216	2.167	<0.001	2.087	2.249
Excess cost (£)	18,051	<0.001	17,192	18,910	18,424	<0.001	17,550	19,297
Prostate								
Cost ratio	1.664	<0.001	1.598	1.732	1.727	<0.001	1.659	1.798
Excess cost (£)	10,726	<0.001	9,924	11,529	11,541	<0.001	10,728	12,354
Head & neck								
Cost ratio	3.250	<0.001	3.016	3.504	3.353	<0.001	3.111	3.615
Excess cost (£)	29,103	<0.001	27,366	30,840	29,972	<0.001	28,127	31,818
Malignant melanoma of skin								
Cost ratio	1.685	<0.001	1.556	1.826	1.783	<0.001	1.642	1.935
Excess cost (£)	7,815	<0.001	6,648	8,982	8,689	<0.001	7,462	9,917
Kidney								
Cost ratio	1.853	<0.001	1.699	2.021	1.872	<0.001	1.721	2.036
Excess cost (£)	12,932	<0.001	11,180	14,684	13,154	<0.001	11,422	14,886
Non-Hodgkin lymphoma								
Cost ratio	3.439	<0.001	3.168	3.732	3.472	<0.001	3.199	3.768
Excess cost (£)	33,808	<0.001	31,394	36,223	34,106	<0.001	31,623	36,589
Oesophagus								
Cost ratio	1.903	<0.001	1.755	2.063	1.918	<0.001	1.767	2.081
Excess cost (£)	14,645	<0.001	12,918	16,372	14,826	<0.001	13,074	16,579
Bladder								
Cost ratio	1.861	<0.001	1.705	2.031	1.881	<0.001	1.732	2.043
Excess cost (£)	15,727	<0.001	13,743	17,712	15,987	<0.001	14,059	17,915
All other cancers								
Cost ratio	2.090	<0.001	2.025	2.157	2.109	<0.001	2.044	2.177
Excess cost (£)	16,020	<0.001	15,298	16,741	16,230	<0.001	15,500	16,959

Notes: CI = confidence interval. Costs are presented undiscounted at 2018 price levels.

Table 4.5.: Full regression models for the multivariable results in Table 4.5 of eight-year excess costs of cancer

	All cancers				Trachea, bronchus and lung				Breast			
Variable	**CR**	***p***	**95% CI**		**CR**	***p***	**95% CI**		**CR**	***p***	**95% CI**	
Cancer	1.985	<0.001	1.956	2.015	1.180	<0.001	1.138	1.223	2.495	<0.001	2.396	2.597
Rural	0.906	<0.001	0.893	0.920	0.889	<0.001	0.858	0.921	0.895	<0.001	0.859	0.932
1 comorbidity	1.285	<0.001	1.210	1.365	1.274	<0.001	1.148	1.415	1.603	<0.001	1.342	1.915
>=2 comorbidities	1.416	<0.001	1.245	1.610	1.346	0.007	1.085	1.669	1.967	<0.001	1.355	2.856
AMI	1.053	0.206	0.972	1.139	1.107	0.209	0.945	1.297	0.948	0.662	0.748	1.203
CHF	0.874	0.002	0.802	0.953	0.759	0.001	0.646	0.893	0.946	0.670	0.734	1.220
PVD	0.939	0.130	0.865	1.019	0.898	0.140	0.779	1.036	0.971	0.825	0.748	1.260
CEVD	0.889	0.004	0.821	0.963	0.851	0.023	0.741	0.979	0.863	0.229	0.678	1.097
Dementia	0.488	<0.001	0.446	0.534	0.514	<0.001	0.435	0.608	0.399	<0.001	0.313	0.510
COPD	0.960	0.226	0.898	1.026	0.845	0.003	0.757	0.943	1.135	0.217	0.928	1.388
Rheumd	1.216	<0.001	1.096	1.349	1.266	0.031	1.022	1.568	1.381	0.024	1.043	1.829
PUD	0.926	0.195	0.824	1.040	0.912	0.479	0.706	1.178	0.670	0.009	0.496	0.906
Diabetes	1.167	<0.001	1.089	1.250	1.036	0.562	0.918	1.169	1.184	0.107	0.964	1.454
Renal	1.082	0.081	0.990	1.183	1.034	0.728	0.857	1.248	1.121	0.358	0.879	1.430
Constant	14,828	<0.001	14,632	15,027	17,627	<0.001	17,098	18,173	11,668	<0.001	11,224	12,129

	Colorectal				Prostate				Head and neck			
Variable	**CR**	***p***	**95% CI**		**CR**	***p***	**95% CI**		**CR**	***p***	**95% CI**	
Cancer	2.167	<0.001	2.087	2.249	1.727	<0.001	1.659	1.798	3.353	<0.001	3.111	3.615
Rural	0.895	<0.001	0.862	0.928	0.867	<0.001	0.833	0.903	0.939	0.097	0.872	1.011
1 comorbidity	1.032	0.730	0.862	1.235	1.758	<0.001	1.518	2.036	1.485	<0.001	1.200	1.837
>=2 comorbidities	0.904	0.575	0.634	1.288	2.209	<0.001	1.625	3.004	1.238	0.331	0.805	1.905
AMI	1.279	0.016	1.047	1.562	0.805	0.022	0.670	0.969	0.963	0.788	0.730	1.269
CHF	1.080	0.489	0.869	1.343	0.804	0.058	0.641	1.008	0.924	0.728	0.592	1.442
PVD	1.237	0.081	0.974	1.571	0.737	0.003	0.604	0.901	0.840	0.317	0.597	1.182
CEVD	0.997	0.979	0.809	1.229	0.903	0.335	0.734	1.111	0.908	0.605	0.630	1.309
Dementia	0.535	<0.001	0.415	0.689	0.381	<0.001	0.299	0.484	0.741	0.313	0.414	1.326
COPD	1.231	0.032	1.017	1.488	0.873	0.111	0.739	1.032	1.112	0.421	0.859	1.438
Rheumd	1.218	0.113	0.955	1.554	0.786	0.122	0.579	1.066	1.691	0.122	0.869	3.290
PUD	1.182	0.310	0.856	1.634	0.726	0.055	0.524	1.007	0.786	0.473	0.407	1.517
Diabetes	1.414	0.001	1.159	1.725	0.992	0.928	0.839	1.173	1.149	0.345	0.862	1.531
Renal	1.304	0.032	1.023	1.662	0.795	0.023	0.652	0.969	1.800	0.018	1.108	2.924
Constant	15,880	<0.001	15,339	16,440	15,744	<0.001	15,147	16,364	12,186	<0.001	11,368	13,062

	Malignant melanomas of skin				Kidney				Non-Hodgkin lymphoma			
Variable	**CR**	***p***	**95% CI**		**CR**	***p***	**95% CI**		**CR**	***p***	**95% CI**	
Cancer	1.783	<0.001	1.642	1.935	1.872	<0.001	1.721	2.036	3.472	<0.001	3.199	3.768
Rural	0.878	0.007	0.800	0.965	0.908	0.029	0.832	0.990	0.928	0.062	0.858	1.004
1 comorbidity	2.196	<0.001	1.505	3.204	1.772	<0.001	1.323	2.372	1.520	0.031	1.038	2.227
>=2 comorbidities	3.998	0.001	1.815	8.804	2.886	0.002	1.491	5.589	2.024	0.069	0.947	4.328
AMI	0.760	0.243	0.479	1.204	1.037	0.849	0.715	1.503	0.797	0.318	0.510	1.244
CHF	0.660	0.127	0.388	1.125	0.579	0.006	0.393	0.852	0.675	0.085	0.431	1.055
PVD	1.113	0.683	0.664	1.866	0.848	0.416	0.569	1.262	0.740	0.248	0.443	1.234
CEVD	0.741	0.189	0.473	1.160	0.810	0.394	0.500	1.314	0.599	0.034	0.373	0.962
Dementia	0.607	0.037	0.380	0.971	0.444	<0.001	0.284	0.693	0.326	<0.001	0.199	0.533
COPD	1.115	0.615	0.730	1.702	0.675	0.024	0.480	0.950	0.983	0.933	0.659	1.467
Rheumd	1.110	0.755	0.576	2.141	0.791	0.393	0.463	1.354	1.168	0.505	0.739	1.846
PUD	1.451	0.076	0.962	2.190	0.587	0.269	0.228	1.510	0.669	0.151	0.387	1.158
Diabetes	0.921	0.689	0.615	1.378	1.012	0.942	0.739	1.385	0.977	0.908	0.661	1.445
Renal	0.721	0.367	0.354	1.467	0.790	0.249	0.529	1.180	0.944	0.806	0.595	1.498
Constant	10,532	<0.001	9,763	11,361	14,449	<0.001	13,402	15,579	13,572	<0.001	12,621	14,594

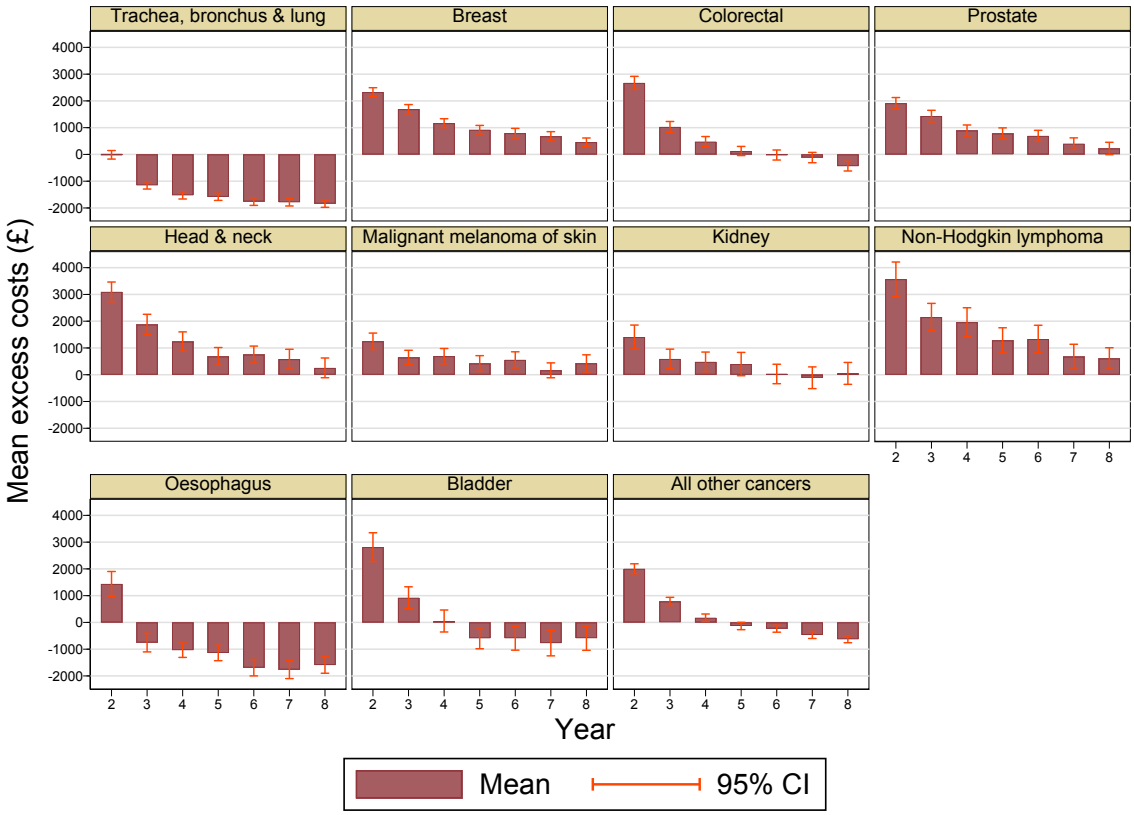
	Oesophagus				Bladder				All other cancers			
Variable	**CR**	***p***	**95% CI**		**CR**	***p***	**95% CI**		**CR**	***p***	**95% CI**	
Cancer	1.918	<0.001	1.767	2.081	1.881	<0.001	1.732	2.043	2.109	<0.001	2.044	2.177
Rural	0.903	0.011	0.835	0.977	0.890	0.003	0.824	0.962	0.916	<0.001	0.888	0.946
1 comorbidity	1.177	0.114	0.962	1.440	1.553	0.024	1.061	2.274	1.021	0.663	0.929	1.122
>=2 comorbidities	1.357	0.200	0.851	2.165	2.586	0.034	1.075	6.224	1.055	0.567	0.879	1.265
AMI	0.913	0.590	0.657	1.270	0.654	0.041	0.436	0.982	1.187	0.021	1.027	1.374
CHF	0.806	0.239	0.563	1.154	0.617	0.039	0.390	0.977	1.019	0.815	0.873	1.188
PVD	0.817	0.361	0.529	1.262	0.749	0.240	0.463	1.213	1.067	0.378	0.924	1.232
CEVD	0.967	0.836	0.704	1.328	0.651	0.036	0.436	0.973	0.964	0.604	0.838	1.109
Dementia	0.446	<0.001	0.324	0.614	0.492	0.006	0.296	0.815	0.524	<0.001	0.447	0.613
COPD	0.946	0.671	0.733	1.221	0.721	0.128	0.473	1.099	1.182	0.003	1.057	1.321
Rheumd	1.068	0.796	0.649	1.758	0.753	0.415	0.382	1.487	1.364	0.001	1.142	1.630
PUD	0.923	0.704	0.610	1.397	0.700	0.249	0.382	1.283	1.051	0.614	0.867	1.273
Diabetes	1.138	0.360	0.862	1.503	0.930	0.747	0.598	1.446	1.309	<0.001	1.168	1.466
Renal	0.853	0.420	0.579	1.256	0.496	0.001	0.325	0.757	1.305	0.001	1.111	1.532
Constant	16,473	<0.001	15,235	17,812	18,506	<0.001	17,074	20,057	14,641	<0.001	14,259	15,034

Notes: CR = cost ratio, CI = confidence interval, AMI = acute myocardial infarction, CHF = congestive heart failure, PVD = peripheral vascular disease, CEVD = cerebral vascular disease, COPD = chronic pulmonary disease, PUD = peptic ulcer. Base levels are not shown for brevity.

4.3.5 Temporal Breakdowns of Excess Costs

Figure 4.15 shows yearly excess costs in the post-diagnosis period. Year 1 excess costs were omitted to show greater detail while maintaining a linear scale, otherwise the confidence intervals would not be visible due to the relatively high levels of year 1 excess costs. These can be seen in Figure 4.16 and were positive for all cancers, as expected, with narrow confidence intervals well above zero. Excess costs turned negative for oesophagus cancer and respiratory cancers after year 2, with costs for respiratory cancers not significantly different from zero at year 2, while those for breast and non-Hodgkin lymphoma remained significantly positive throughout the eight-year follow-up. Prostate, skin, and head and neck cancers also showed positive mean costs in all years, although not significantly different from zero in year 7 for skin and year 8 for prostate and head and neck.

Figure 4.15.: Trajectories of excess costs by year after year 1 and by cancer type



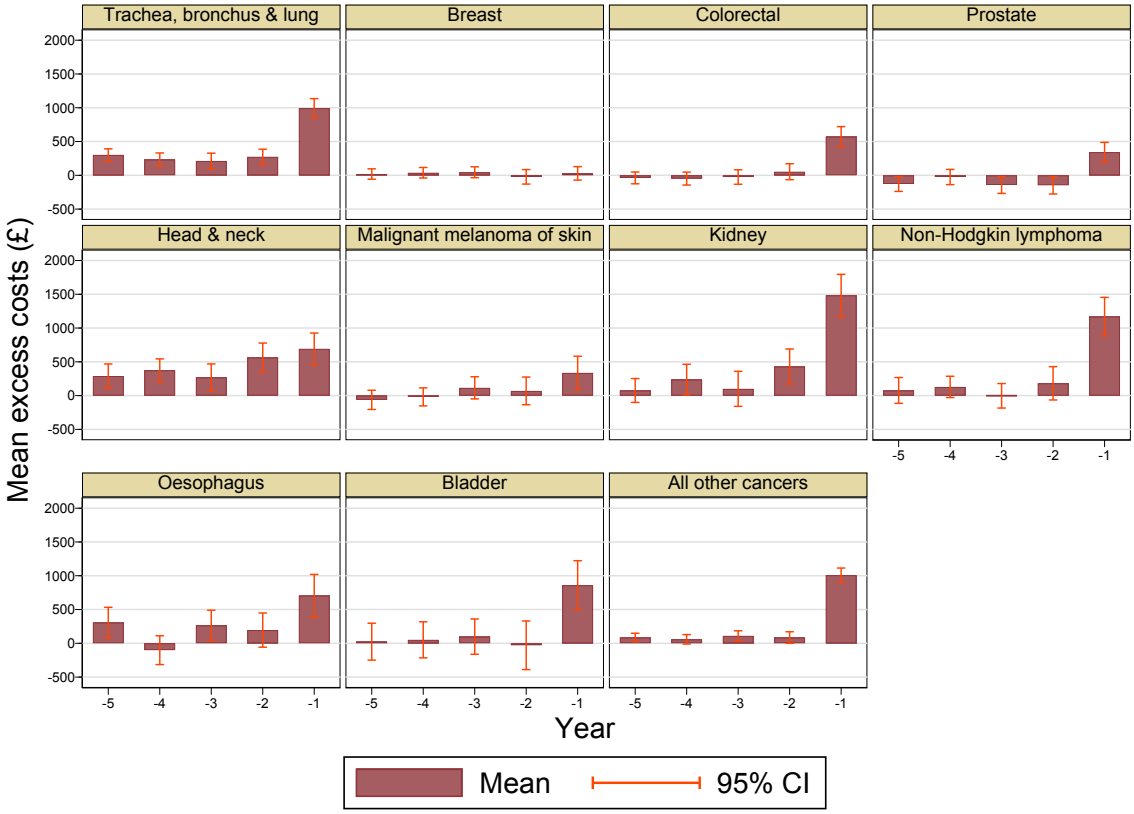
Notes: CI = confidence interval. Costs are presented undiscounted at 2018 price levels. Years are relative to the diagnosis-event. Year 1 is presented separately due to relatively high excess costs.

Figure 4.16.: Excess costs in year 1 by cancer type

Note: CI = confidence interval

Excess costs shown in Figure 4.17 were positive and substantial for all pre-diagnosis years for respiratory cancers and for head and neck cancers, but were not significantly different from zero for breast cancer in all years. Other cancers showed significantly positive excess costs only in the year directly before the diagnosis-event with costs in the approximate region of £500 to £1,000.

Figure 4.17.: Trajectories of excess costs by year in the pre-diagnosis period and by cancer type



Notes: CI = confidence interval. Costs are presented undiscounted at 2018 price levels. Pre-diagnosis costs excluded prescriptions

Yearly costs for all years are given in tabular form in Table 4.6 to allow the calculation of costs with custom discount rates and time periods, which may include pre-diagnosis costs if desired. Phase-of-care costs are shown in Table 4.7 with mostly positive excess costs observed in the post-diagnosis phases, although the confidence intervals in some phases were wide for cancers with high mortality.

Table 4.6.: Estimates of excess costs by year and cancer type

Trachea, bronchus and lung				Breast				Colorectal			
Year	Mean (£)	95% CI (£)		Year	Mean (£)	95% CI (£)		Year	Mean (£)	95% CI (£)	
-5	303	214	392	-5	19	-58	96	-5	-38	-126	49
-4	234	138	331	-4	38	-40	116	-4	-48	-144	47
-3	213	98	328	-3	45	-36	126	-3	-25	-133	82
-2	275	163	387	-2	-22	-130	85	-2	53	-65	171
-1	993	851	1,135	-1	29	-71	129	-1	573	425	721
1	12,811	12,519	13,103	1	8,898	8,675	9,122	1	14,319	13,985	14,653
2	-14	-174	145	2	2,329	2,165	2,493	2	2,674	2,430	2,918
3	-1,157	-1,293	-1,020	3	1,691	1,519	1,863	3	1,028	826	1,231
4	-1,529	-1,666	-1,392	4	1,168	1,003	1,334	4	479	290	668
5	-1,586	-1,718	-1,454	5	924	764	1,084	5	128	-42	299
6	-1,763	-1,904	-1,621	6	795	616	974	6	-22	-210	165
7	-1,790	-1,927	-1,653	7	677	504	850	7	-117	-310	76
8	-1,840	-1,977	-1,703	8	459	304	615	8	-440	-619	-261
Sum (1-8)	3,133			Sum (1-8)	16,942			Sum (1-8)	18,051		
Sum (all years)	5,150			Sum (all years)	17,051			Sum (all years)	18,565		
Prostate				**Head & neck**				**Malignant melanoma of skin**			
Year	Mean (£)	95% CI (£)		Year	Mean (£)	95% CI (£)		Year	Mean (£)	95% CI (£)	
-5	-128	-238	-18	-5	290	112	468	-5	-63	-205	79
-4	-25	-137	88	-4	375	206	544	-4	-19	-153	115
-3	-143	-268	-17	-3	270	72	469	-3	115	-48	279
-2	-148	-276	-19	-2	563	349	778	-2	69	-135	274
-1	343	199	488	-1	693	459	926	-1	337	91	583
1	4,381	4,155	4,606	1	20,617	19,832	21,401	1	3,663	3,333	3,993
2	1,913	1,700	2,126	2	3,097	2,730	3,464	2	1,249	943	1,555
3	1,434	1,218	1,649	3	1,875	1,493	2,257	3	647	381	913
4	895	689	1,100	4	1,239	879	1,600	4	693	407	979
5	784	576	992	5	684	352	1,015	5	419	125	712
6	694	488	900	6	754	437	1,070	6	556	255	856
7	404	189	618	7	582	212	951	7	165	-114	443
8	223	-7	453	8	256	-113	625	8	424	104	744
Sum (1-8)	10,726			Sum (1-8)	29,103			Sum (1-8)	7,815		
Sum (all years)	10,627			Sum (all years)	31,294			Sum (all years)	8,255		
Kidney				**Non-Hodgkin lymphoma**				**Oesophagus**			
Year	Mean (£)	95% CI (£)		Year	Mean (£)	95% CI (£)		Year	Mean (£)	95% CI (£)	
-5	76	-101	252	-5	77	-114	268	-5	308	84	532
-4	237	11	464	-4	129	-29	286	-4	-102	-315	111
-3	100	-159	360	-3	-3	-184	178	-3	266	42	490
-2	431	172	690	-2	182	-64	429	-2	195	-58	448
-1	1,484	1,173	1,795	-1	1,170	886	1,454	-1	706	392	1,019
1	10,092	9,500	10,683	1	22,221	21,025	23,418	1	21,194	20,226	22,162
2	1,408	961	1,855	2	3,569	2,928	4,211	2	1,439	976	1,902
3	589	224	954	3	2,154	1,642	2,665	3	-755	-1,101	-409
4	481	115	846	4	1,962	1,423	2,502	4	-1,039	-1,307	-771
5	398	-37	834	5	1,284	814	1,754	5	-1,143	-1,430	-855
6	28	-335	392	6	1,334	821	1,847	6	-1,691	-1,998	-1,385
7	-114	-519	292	7	675	211	1,138	7	-1,766	-2,099	-1,433
8	49	-359	458	8	609	210	1,008	8	-1,594	-1,899	-1,288
Sum (1-8)	12,932			Sum (1-8)	33,808			Sum (1-8)	14,645		
Sum (all years)	15,260			Sum (all years)	35,363			Sum (all years)	16,018		
Bladder				**All other cancers**				**All cancers**			
Year	Mean (£)	95% CI (£)		Year	Mean (£)	95% CI (£)		Year	Mean (£)	95% CI (£)	
-5	24	-249	297	-5	86	23	149	-5	80	47	113
-4	51	-215	318	-4	59	-11	128	-4	74	39	109
-3	99	-163	360	-3	111	36	185	-3	80	41	119
-2	-29	-389	330	-2	85	0	171	-2	103	60	147
-1	859	497	1,222	-1	1,008	903	1,114	-1	707	656	758
1	14,489	13,693	15,285	1	14,525	14,223	14,827	1	12,506	12,372	12,639
2	2,814	2,278	3,350	2	2,005	1,820	2,190	2	1,853	1,772	1,934
3	922	514	1,331	3	785	632	937	3	735	664	805
4	53	-357	463	4	176	36	315	4	246	179	313
5	-594	-984	-203	5	-133	-271	4	5	-14	-79	51
6	-592	-1,036	-149	6	-235	-365	-105	6	-134	-200	-68
7	-775	-1,249	-301	7	-470	-601	-340	7	-315	-381	-248
8	-590	-1,043	-138	8	-632	-759	-505	8	-450	-514	-385
Sum (1-8)	15,727			Sum (1-8)	16,020			Sum (1-8)	14,427		
Sum (all years)	16,732			Sum (all years)	17,369			Sum (all years)	15,471		

Notes: CI = confidence interval. Estimates measured as mean differences between cancer and non-cancer cohorts. Years are relative to the diagnosis-event. Costs are presented undiscounted at 2018 price levels.

Table 4.7.: GLM regression estimates of excess costs by phase-of-care

Cancer type	Phase	AME (£)	*p*	95% CI (£)		Cancer type	Phase	AME (£)	*p*	95% CI (£)	
Trachea, bronchus and lung	pre-diagnosis	1,598	<0.001	1,220	1,976	**Breast**	pre-diagnosis	-206	0.2	-521	109
	initial	11,153	<0.001	10,730	11,575		initial	11,148	<0.001	10,754	11,542
	continuing	1,718	0.001	713	2,723		continuing	6,279	<0.001	5,649	6,909
	end-of-life	1,413	<0.001	819	2,006		end-of-life	5,540	<0.001	4,460	6,621
	sum	15,882					sum	22,761			
Colorectal	pre-diagnosis	-27	0.891	-408	355	**Prostate**	pre-diagnosis	-835	<0.001	-1,240	-429
	initial	17,151	<0.001	16,567	17,735		initial	3,752	<0.001	3,504	4,000
	continuing	6,178	<0.001	5,430	6,926		continuing	4,414	<0.001	3,701	5,127
	end-of-life	5,842	<0.001	4,972	6,713		end-of-life	4,757	<0.001	3,885	5,630
	sum	29,144					sum	12,089			
Head & Neck	pre-diagnosis	1,462	<0.001	749	2,174	**Malignant melanoma of skin**	pre-diagnosis	-199	0.531	-822	424
	initial	26,505	<0.001	24,729	28,280		initial	3,580	<0.001	3,150	4,011
	continuing	10,687	<0.001	9,015	12,359		continuing	2,135	<0.001	1,127	3,143
	end-of-life	8,756	<0.001	6,606	10,906		end-of-life	5,048	<0.001	3,239	6,856
	sum	47,410					sum	10,564			
Kidney	pre-diagnosis	1,663	<0.001	798	2,527	**Non-Hodgkin lymphoma**	pre-diagnosis	856	0.023	118	1,593
	initial	9,325	<0.001	8,430	10,220		initial	28,732	<0.001	26,378	31,087
	continuing	6,015	<0.001	4,307	7,722		continuing	13,675	<0.001	11,594	15,757
	end-of-life	4,812	<0.001	3,192	6,433		end-of-life	16,904	<0.001	13,667	20,141
	sum	21,815					sum	60,167			
Oesophagus	pre-diagnosis	929	0.025	114	1,744	**Bladder**	pre-diagnosis	439	0.417	-621	1,498
	initial	20,107	<0.001	18,403	21,811		initial	13,660	<0.001	12,446	14,875
	continuing	1,429	0.152	-525	3,382		continuing	8,157	<0.001	6,262	10,053
	end-of-life	8,736	<0.001	6,702	10,770		end-of-life	6,018	<0.001	4,347	7,688
	sum	31,201					sum	28,274			
All other cancers	pre-diagnosis	816	<0.001	540	1,091	**All cancers**	pre-diagnosis	525	<0.001	386	664
	initial	15,643	<0.001	15,105	16,180		initial	12,756	<0.001	12,541	12,971
	continuing	8,508	<0.001	7,767	9,250		continuing	6,047	<0.001	5,752	6,343
	end-of-life	6,812	<0.001	6,131	7,492		end-of-life	5,388	<0.001	5,062	5,714
	sum	31,778					sum	24,716			

Notes: GLM = generalised linear model, CI = confidence interval, AME = average marginal effect. Estimates were obtained as AMEs from GLM models. Phases are relative to the diagnosis-event. Costs are presented undiscounted at 2018 price levels.

4.4 Discussion

4.4.1 Key Results

Previous studies have found substantial excess costs for common cancers [105, 177, 178, 135, 179, 21, 180, 181, 182, 185, 186] with considerable variation across cancer sites [179, 105, 186]. This analysis confirmed and extended these findings for the Scottish population using patient-level data while measuring costs for cancers that have previously received little study, such as bladder cancer and kidney cancer. Considerable variation was found between cancer sites, with the highest eight-year excess costs measured for non-Hodgkin lymphoma at £34,106, and the lowest in cancers of the trachea, bronchus and lung at £3,153. Analysis of annual costs reflected the influence of survival, with excess costs for high-mortality cancers such as lung cancer and bladder cancer turning negative two to three years after diagnosis,

leading to relatively low eight-year costs for cancers with high mortality. However, cancers with low mortality, such as skin cancer, also accrued low excess costs despite costs remaining elevated. Higher eight-year costs were found for cancers with moderate survival where costs remained elevated throughout eight-year follow-up. When costs for survivors were considered using a phase-of-care approach, positive excess costs were observed during all post-diagnosis phases in all cancers studied, although this was not significant for cancer of the oesophagus in the continuing phase.

4.4.2 Interpretation

Cost Patterns Over Time

Care must be taken when comparing post-diagnosis event costs with those in the pre-diagnosis period, particularly when considering yearly periods. This is partly because mortality-related attrition operating in the post-diagnosis periods did not apply pre-diagnosis (as all patients were required to be alive at baseline) and also because prescription costs were not included pre-diagnosis. Hence, costs in the pre-diagnosis and post-diagnosis periods were not aggregated, in line with the study objective of examining excess costs in the years following diagnosis. However, pre-diagnosis costs were noted as a risk factor in Section 3.3, and the costs shown in Figure 4.17 suggest that excess costs were positive during the pre-diagnosis period, implying that post-diagnosis costs may not have captured the full magnitude of resources used. Costs were substantially higher for patients of all cancer types than their non-cancer controls during the initial period after diagnosis. This was expected on account of the greatly increased use of healthcare during the treatment period. Costs remained elevated for all cancer types during the continuing period. Contributors to these positive excess costs may have been ongoing surveillance, complications and new cancers [7]. Costs were also elevated during the end-of-life period, which for some patients may have comprised all of their post-diagnosis trajectories.

Cost Patterns Across Cancer Types

Explanations for the patterns of excess costs across cancers may reside in how the tumour characteristics affected the cost trajectories. Cancers with very high mortality are likely to have had higher disease severity, resulting in considerable resource use over a short period, and were likely to have been detected at later stage, with high amounts of palliative care in the end-of-life phase. Fewer survivors lead to lower mean resource use in the years after diagnosis, thus lowering excess costs over longer periods. It

seems likely that cancers with very high mortality rates would have had negative overall excess costs over longer periods than those shown in this study, particularly if additional services such as social care and mental health care were accounted for. In contrast, cancers with relatively low mortality, such as skin cancer, used less hospital care in the short term, but continued to accrue moderate levels of healthcare over the longer term. Cancers with moderate survival also accrued substantial resource use in the short term, and levels of resource use remained elevated for many years after diagnosis, leading to higher costs. This explanation would concur with Blakely et al. (2015) [179], who found that moderate survival was associated with the highest costs, while both low and high survival cancers were associated with lower costs.

Discounting

Costs were reported undiscounted to highlight that they represented past expenditures rather than the present value of projected future costs. The reasons for presenting undiscounted costs were described in Section 3.2 and Section 2.3.6. As a high proportion of costs were observed in the first year after diagnosis, with costs declining sharply thereafter, the difference between discounted and undiscounted costs is unlikely to have been substantial, and in the case of cancers with high mortality could have led to lower excess cost estimates for undiscounted costs. For the purposes of evaluations and other costings where discounting may be appropriate, discount rates can be applied to yearly costs presented in Table 4.6. Adjustment for comorbidities increased the estimates of excess costs, despite comorbidities being higher in the cancer cohort. This may have resulted from higher mortality rates among cancer patients lowering the costs for cancer patients with comorbidities, compared to patients with comorbidities and without cancer. It is possible the effect was even greater than observed, because the episode-based costing method would underestimate costs for patients with very long stays, as discussed in Section 3.4.

Comparison to Other Studies

The excess costs for lung-related cancers in my analysis were notably lower than those reported in other studies [177, 135, 21]. An extreme example is Kutikova et al. (2005) [177] who reported costs for US lung cancer patients 15.79 times that of similar controls (mean excess cost £44,168), compared to 1.18 times in my study. The lower cost ratio in my analysis can be explained by the effects of survival over the longer follow-up in my study. Kutikova et al. (2005) [177] used a two-year follow-up while I used eight years. In my analysis Figure 4.4 shows that costs in the year after diagnosis

were approximately eight times higher for lung cancer patients than those of controls. Excess costs shown in Figure 4.15 were not significantly different in year 2 and thereafter negative, reducing the cost ratio at each until the end of the eight-year follow-up. The shorter follow-up may seem reasonable because survival at two years is very low for lung cancer patients. However, it may give a misleading impression of the overall long-term societal costs, as well as a misleading impression of the economic effects of reducing lung cancer incidence. Over a longer time frame the societal costs are considerably lower. Other studies with longer follow-ups also found lower excess costs for lung cancer, such as Blakely et al. (2015) (mean excess cost £15,623) in New Zealand [179] and Yabroff et al. (2008) (mean excess cost for men £33,977, women £35,503) in the US [21] both with five-year follow-ups. These studies did not report costs for controls so a cost ratio was not available, however that of Banegas et al. (2018) [135] was 1.47 for lung cancer, which is considerably closer to my figure than to that of Kutikova et al. (2005) [177]. It should be noted that Yabroff et al. (2008) [21] used probabilistic methods based on survival to estimate costs rather than following up all patients over the full five years, and only studied patients over 65 years old, who are believed to have lower costs than younger patients [135, 105]. The relatively high excess costs for lung cancer in US studies may have been a consequence of high drug prices for palliative treatments that are not considered cost-effective in other countries [44].

My primary approach assigned zero costs to deceased individuals in each remaining year of follow-up. Consequently, cancers with very high mortality such as lung cancer and bladder cancer were dominated by zero counts in later years, leading to negative excess costs which lowered the eight-year costs. Using a similar approach, Blakely et al. (2015) [179] also found lower costs associated with high-mortality cancers such as lung cancer (mean excess cost £15,623), and pancreas cancer (mean excess cost £12,390) while higher costs were associated with moderate mortality cancers such as leukaemia (mean excess cost £51,177) and non-Hodgkin lymphoma (mean excess cost £38,787). The high excess costs associated with non-Hodgkin lymphoma were also found in Yabroff et al. (2008) (mean excess cost for men £41,464, women £39,177) [21]. Relatively moderate excess costs for breast cancer in my analysis were in agreement with other studies [180, 182, 105, 21, 179].

4.4.3 Strengths and Limitations

The large sample size allowed me to compare costs with high statistical power for the 10 most common cancers in Scotland and for all other cancers combined. I was able

to measure costs in excess of individuals matched precisely on age, sex, NHS Health Board and SIMD quintile, over a long time frame. The combination of annual costs and phase-of-care costs allowed me to study how excess costs evolved for the entire cohort and for survivors. However, linked administrative data bring limitations in the range of available variables as described in Section 3.4. Data on in-hospital drugs, primary care, social care, and mental health were lacking. It is not certain how this would affect results but cancer patients have been reported to make greater use of mental health services [175], suggesting underestimation of excess costs. However, if total costs were increased, the impact of cancer treatment may become a relatively smaller portion of total costs, lowering the cost ratio. Additionally, the poorer survival of cancer patients could reduce the absolute magnitude of excess costs. The datasets by their nature excluded private healthcare, meaning that full healthcare use was not captured. It is not clear that this would bias excess costs downward, as the very high cost of cancer care could reduce its accessibility to private care, while lower-cost private care may be utilised more by individuals without life-threatening disease. Some cancers may also be more amenable to private care than others [179], meaning this bias could affect some cancers more than others. An additional limitation was that the analysis of multiple cancers required a broad approach with inevitable simplification of the complexities of individual cancers. However, including multiple cancers within a single study allowed comparison across cancer sites within a single population, which is arguably a more useful approach due to difficulties in generalising results described in Section 4.4.4.

4.4.4 Generalisability

The issues related to generalising the results of Chapter 3, described in Section 3.4 also apply here. The findings will be more relevant to countries with similar levels of economic development, general health and with public healthcare systems. Findings relating to variation between cancer sites and how costs change over time are more likely to generalise than the absolute magnitudes of costs, which are closely intertwined with the specifics of NHS Scotland healthcare systems and the methods of cost measurement used in this study.

4.4.5 Policy Implications and Future Research

Cancer Types

The high excess costs of breast cancer and colorectal cancer, combined with their high prevalence, suggest that resources invested in reducing incidence of these cancers may be efficiently targeted. Lung-related cancers had much lower excess costs, even in aggregate, suggesting that while the focus on reducing incidence of these cancers may be effective in preventing premature deaths, the overall long-term cost savings to health systems may not be very substantial. Other cancers with lower prevalence such as bladder cancer and non-Hodgkin lymphoma had considerably higher excess costs, both per-person and in aggregate. This suggests there may more be potential for reducing costs in less common cancers than in trachea, bronchus and lung cancers.

Survival and Costs

The low excess costs for high mortality cancers reinforces the relationship between survival and costs found in Chapter 3, which suggests that improvements to survival are likely to entail higher costs. In particular, if the prevalence of high mortality cancers falls, while that of cancers with moderate survival and higher economic costs rises, and the prevalence of chronic diseases also rises, healthcare costs are likely to increase. The additional years of life gained are also likely to lead to higher social care costs, higher pension costs, and greater demand for housing and transport.

Prevention

It is not clear that prevention will ultimately reduce costs as the costs may simply be delayed rather than avoided, and the problem of how to encourage populations to alter their lifestyles would need addressed. Policymakers should carefully consider whether the benefits of interventions are justified. Further study into the wider economic effects of increased longevity and of intertemporal preferences could inform better policymaking in this area.

End of Life Costs

My results suggested that a major proportion of costs for cancer patients are accrued in the end-of-life period. However, when compared to patients without cancer, the excess costs in this phase were not very substantial. As everyone must ultimately die, it is not clear whether end-of-life costs could be avoided or merely postponed. This

indicates that studies which measure cancer costs using an incident method over short time frames may overestimate cancer costs, and that studies with costs discounted over longer time frames may also overestimate cancer costs. However, if costs in the end-of-life period are not cancer-specific, the reduction of such costs could bring greater cost savings than reducing them solely in cancer patients.

Covid and Other Factors

The impact of Covid has put considerable pressure on the NHS, leading to delays in the screening of many cancers [49], which can impact the stage at which cancers are detected and alter the trends in incidence and prevalence [8]. The effect on healthcare costs will be an important topic for future research. Other future investigations could focus on the impact of cancer on social care use and mental health use. Another important task will be untangling the relationship between cancer costs, pre-existing long-term conditions, and their common underlying risk factors such as diet. Further questions remain around to what extent costs are determined by the differing characteristics of specific cancer types, and more general clinical and patient factors. However, policymakers should be cautious when interpreting costing studies, as differing methodologies, time frames and other factors can make substantial differences to final costs. The uncensored approach of this study was appropriate for estimating healthcare costs to a national public health service, while different approaches may be more relevant to other applications and healthcare systems.

4.4.6 Conclusion

This analysis extended the results of Chapter 3 by comparing healthcare use and associated costs with a matched control group. I charted trajectories of healthcare use and associated costs for people with and without cancer over eight years to better understand differences in trends over time. Confounder-adjusted excess costs associated with cancer were measured over the eight-year period for the 10 most common cancers in Scotland and other cancers combined. I also analysed annual excess costs to enhance the understanding of how they developed over time. This information provides insights into the long-term costs of cancer that will be of benefit to health professionals, health economists and policymakers. A further objective of the thesis was to examine how the presence of long-term conditions (LTCs) influenced costs for patients after a cancer diagnosis. Chapter 5 will describe how I utilised the dataset used in chapters 3 and 4 to meet this objective.

5 Inpatient Costs for Patients with Long-Term Conditions

5.1 Introduction

In chapters 3 and 4, I estimated the healthcare use and associated costs of cancer patients, and compared them with those of a control group without cancer to estimate excess costs. Trajectories of costs were described and cost drivers analysed. Regression analysis indicated that comorbidities were associated with healthcare costs. However, the direction and magnitude of association varied by condition. Heterogeneous survival across different underlying conditions is a possible explanation, while patient characteristics such as frailty may also have contributed. The extent to which costs could be attributed to cancer, to underlying long-term conditions (LTCs), and to common risk factors was unclear, suggesting that further investigation of the relationship between cancer and underlying health problems would improve understanding of healthcare use and associated costs.

Before proceeding is a note on terminology. In this study the terms LTCs and comorbidities will often be used synonymously. However, for the analytical part of the analysis, LTCs refer to the specific conditions analysed while comorbidities refer to any conditions other than cancer, which may include the LTCs being analysed but could also include other conditions.

5.1.1 Background Rationale

Before the Covid epidemic, LTCs were described as *"the main challenge facing healthcare systems worldwide"* [188] and they accounted for the majority of high-income countries' health spending [151]. Their rising prevalence has slowed or reversed health improvements in high-income countries, with healthy life expectancy falling behind life expectancy meaning people are living longer in poor health [189]. In 2019, the five main causes of disability-adjusted life years (DALYs) worldwide were ischaemic heart disease, stroke, diabetes, chronic obstructive pulmonary disease (COPD) and lung cancer for people aged 50–74 years old, while for people aged 75 and older they were ischaemic heart disease, stroke, COPD, Alzheimer's and diabetes [190]. Combined, cardiovascular disease (CVD), diabetes and cancer account for around two-thirds of all US deaths [191] while CVDs were the leading cause of death globally in 2016 causing 31% of all deaths [192]. Although healthcare has more recently focused on Covid, the impact of LTCs has not lessened. Covid has interacted with chronic diseases and their risk factors, such as obesity, high blood pressure and air pollution to exacerbate Covid mortality, leading to calls for government action to address the syndemic [189]. LTCs have been a major factor driving Covid mortality and

will likely be fundamental in driving health and policy in the foreseeable future [189].

Cancer and LTCs tend to be intertwined [48, 193, 192] due to common risk factors such as age [188, 194], obesity [195, 196, 197, 192], smoking [192], alcohol [33, 193, 192], inactivity [192, 198], diet [199], and socioeconomic status [188, 200]. In western Europe and high-income North America, tobacco was the highest risk factor in 2019 but declining, while the next four biggest risks high body mass index (BMI), dietary risks, high fasting plasma glucose and high systolic blood pressure were all increasing in prevalence [201]. Cancer survivors are more likely to have LTCs than adults without cancer, with four or more chronic conditions present in over 10% of patients [6]. While cancer survival has generally been improving, this may increase the costs associated with comorbidities [6]. This is highly relevant to the UK where demographic and lifestyle changes have been associated with high levels of poor health, leading to increasingly complex health and social care needs [62]. The Scottish population in particular is believed to have poor health [61] with 44% of adults living with an LTC, and life expectancies at 77.1 for men and 81.8 for women lower than the western European averages of 78.9 for men, 83.7 for women [16]. The increase in prevalence of LTCs caused by these risk factors entails substantial economic costs. Healthcare related to LTCs accounted for 70% of UK healthcare spending and rising in 2016 [16]. The increase of 11 million obese people between 2011–2030 is estimated to cost the NHS an additional £1.9–2 billion per year through increased incidence of diabetes, CVD and cancer [195]. Physical inactivity is estimated to have cost NHS Scotland £94 million during 2010–2011 due to an increased risk of diabetes, CVD and cancer [198].

Higher comorbidity has been found associated with higher excess costs [202], with multiple conditions believed to be superadditive, i.e. the association with multiple conditions was higher than the expected sum. The superadditive association was found to be stronger in younger adults [151], however, where one of the conditions was cancer, the excess costs were not significantly different from zero [151]. Another study found that total costs for cancer survivors increased with a higher number of comorbidities [148], while a different study found little association between the cost of chronic conditions and cancer [203]. These results suggest a complex relationship between cancer and other conditions and it is not clear how cancer would affect costs for patients with LTCs. While LTCs lead to higher healthcare use, and a cancer diagnosis will likely increase it further, over the long term the higher mortality rates of cancer patients may reduce healthcare use overall. Another possibility is that patients with LTCs receive less aggressive treatment than healthier patients due to greater sensitivity to the side effects of treatment. Measurement of cancer costs for patients

with LTCs could help to untangle these issues and improve understanding of the costs of cancer, and how they may develop in future as the prevalence of LTCs rises.

5.1.2 Aims and Objectives

Although the presence of comorbidities may increase costs for some index conditions, the impact of cancer on the costs for patients with pre-existing LTCs has received little study, despite the interconnection of common risk factors. This analysis aimed to fill this knowledge gap while bringing a greater understanding of cancer costs in the Scottish population. The overall aim was to better understand the costs of cancer for patients with underlying LTCs. To meet this aim I posed the following research questions.

1. How does a cancer diagnosis affect the healthcare costs of individuals with pre-existing LTCs?
2. How does a cancer diagnosis affect costs specific to the LTCs?
3. How do the costs change over time?

To answer these questions, the following objectives were specified.

1. Measure and chart cost trajectories of cancer patients with pre-existing LTCs, and similar individuals with the same LTCs but no cancer diagnosis.
2. Measure the association of a cancer diagnosis with healthcare costs for patients with LTCs.
3. Measure the association of a cancer diagnosis with costs specific to the LTCs.

5.2 Methods

5.2.1 Study Overview

This was a retrospective cohort study using linked administrative data to measure incidence costs of healthcare use for cancer patients with pre-existing LTCs. The perspective was that of a public healthcare provider, which can be considered a societal one. The setting was NHS Scotland regions as described in Section 3.2. A five-year exposure window for the cancer diagnosis was chosen, spanning 1 January 2006 to 31 December 2010. This window was longer than in previous analyses to maximise numbers of patients in LTC cohorts and to avoid disclosure issues, which could arise through differencing of numbers across this analysis and previous ones. As

in previous analyses I used a follow-up time of eight years from the exposure event for each individual. Records from 1996 onward and prior to the exposure event were used to gain information on comorbidities including those of the LTCs under study, making the full study period 1 January 1996 to 31 December 2018.

5.2.2 Participants

Exposure was a first cancer diagnosis during the exposure period, recorded as an entry in the SMR06 database with one of the ICD10 codes listed in Section 3.2.2. Controls were individuals who received no cancer diagnosis at any time during the study time frame. All controls were taken from a dataset of individuals previously matched to cancer patients as described in Section 4.2. Precise matching of cancer patients to controls was not performed in this analysis as the overlapping sets of cancer patients and controls would have been low in numbers, therefore lacking statistical power and potentially causing privacy issues around the disclosure of low numbers. All non-cancer individuals were assigned a pseudo-diagnosis event by eDRIS as described in Section 3.2.3. All individuals were at least 18 years old at the time of the diagnosis event.

Three LTC groups: CPD, CVD and diabetes were defined based on the comorbidity codes of Quan et al. (2005) [204] below, plus a no-comorbidity group with zero recorded comorbidities prior to the diagnosis event. The ICD codes were derived from the main condition variable in SMR01 entries at any time between January 1st 1996 and the diagnosis event. An individual could belong to multiple LTC groups, except for individuals in the no-comorbidity group. The ICD10 codings were:

CPD: J40, J41, J42, J43, J44, J45, J46, J47, J60, J61, J62, J63, J64, J65, J66, J67, I278 , I279, J684, J701, J703

CVD: I21, I22, I252, I43, I50, I099 , I130, I132, I255, I420, I425, I426, I427, I428, I429, P290, I70, I71, I731, I738, I739, I771, I790, I792, K551, K558, K559, Z958, Z959, G45, G46, I60, I61, I62, I63, I64, I65, I66 , I67, I68, I69, H340

Diabetes: E100, E101, E106, E108, E109, E110, E111, E116, E118, E119, E120, E121, E126, E128, E129, E130, E131, E136, E138, E102, E103, E104, E105, E107, E112, E113, E114, E115, E117, E122, E123, E124, E125, E127, E132, E133, E134, E135, E137, E142, E143, E144, E145, E147

No-comorbidity: No recorded comorbidity of any type prior to the diagnosis-event, i.e. zero SMR01 records containing the comorbidities defined by Quan et al. (2005)

[204] from the beginning of the study period to the diagnosis event.

Participants could belong to more than one LTC group, though individuals in the no-comorbidity group could not be included in an LTC group. Rather than include groups for multiple LTCs, each additional LTC was treated as a comorbidity to the LTC under investigation in a particular analysis. The adjustment for comorbidities is described in Section 5.2.4.

5.2.3 Data Sources

Datasets, access and linkage were the same as those described in Section 3.2.3, except that I excluded SMR00 (outpatient) and PIS (prescriptions) datasets. I chose to exclude these datasets because no reliable method of assigning costs to specific LTCs was available. Previous analyses indicated that these were minor contributors to overall costs, while entailing high computational burdens. Therefore only inpatient and daycase episodes derived from SMR01 were included in cost estimates. Completeness of SMR01 records has been reported at 99–100% in all years during the period 2016–18 [205].

The cohorts consisted of individuals with cancer and non-cancer individuals who had previously been matched to similar cancer patients. Hence the sample was non-random and the characteristics of patients with LTCs may not have been typical of the Scottish population. To minimise bias, I adjusted for multiple confounders as will be described in Section 5.2.5.

As discussed in Section 3.4, comorbidities in SMR01 are known to be under-reported [161], which is a known issue with administration data [118]. Attribution of comorbidities is also known to be imperfect; clinical coding accuracy of the main condition and main operation in SMR01 was found to be 88%–90% for all conditions between 1992 and 2015 [161]. However, for common conditions accuracy was higher at 96.3% with under-recording 6.4 times greater than over-recording. Diabetes was reported as 97.2% accurate, COPD 99.5%, ischaemic heart disease 97.1%, myocardial infarction 98.6% [161]. Under-reporting was 13.7% for diabetes, 14.7% for COPD, 19.3% for ischaemic heart disease, and 6.8% for myocardial infarction [161].

5.2.4 Variables

Outcomes

Two categories of costs were defined: total and LTC-specific. Total costs were defined as all inpatient/daycase costs recorded in SMR01 records in a particular period. LTC-specific costs were defined as the inpatient/daycase costs attributed to a particular LTC, with attribution derived from the main condition recorded in the SMR01 episode. For all costs I used a per-episode method of costing as described in Section 3.2.4. To determine the level of association between cancer and costs, excess costs for both total and LTC-specific costs were calculated using methods described in Section 3.2.4. An additional outcome was survival time, which was determined using date of death as described in Section 3.2.5. Survival was a secondary outcome that was investigated to enhance understanding of how the development of costs over time related to mortality.

Exposure

Exposure for all LTC groups was defined as a cancer diagnosis recorded in the SMR06 dataset. Non-cancer patients were assigned a pseudo-diagnosis event time concurrent with that of a cancer patient (though not necessarily one in this analysis) as described in Section 3.2.5.

Confounders

The confounders, age, sex, SIMD, rurality, diagnosis-event year, and individual comorbidities were included in multivariable regression. These were determined from literature, expert opinion and analyses in previous chapters. Participants were taken from a larger matched sample, therefore similar in terms of age, sex, SIMD and geographical region. As model parsimony was not a priority a non-parsimonious approach to variable selection was taken, albeit the included list of confounders was limited by the availability and quality of variables in the dataset, a common problem with datasets not created for purposes of analysis [118]. Health board regions were aggregated into network regions to minimise low numbers that could have caused problems with the disclosure of sensitive patient information.

Age: Inspection of scatter plots indicated a complex relationship between age and costs, that varied across LTC groups. Model fitness tests using fractional polynomials did not improve model fit compared to the five age categories used in descriptive

statistics and had little effect on the cancer coefficient, which was of most interest, while complicating model interpretation. Therefore I chose to use the simpler model using age categories.

Year of event: As the exposure window of five years was long enough for secular effects to be substantial, I included a year of event variable to adjust for any biases across exposure cohorts. In the cancer group this referred to the year in which the diagnosis occurred (i.e. the date recorded in SMR06). The non-cancer group was assigned a pseudo-diagnosis event date as described in Section 4.2.3. The year of event date was the calendar year in which the diagnosis-event dates occurred.

Comorbidities: As in previous analyses, I used binary variables for comorbidities in adjusted models. Numbers for the comorbidities hemiplegia, liver diseases and human immunodeficiency virus (HIV) were low in particular LTC groups, which could have caused low statistical power and disclosure issues in reporting. These variables were aggregated to an *other comorbidity* variable. For the same reason I incorporated diabetes complications into a single diabetes variable covering the diabetes LTC group. The full list of comorbidities was acute myocardial infarction, congestive heart failure, peripheral vascular disease, cerebral vascular disease, dementia, chronic pulmonary disease, rheumatoid disease, peptic ulcer, diabetes, renal disease, other comorbidity. Only comorbidities outside the LTC group under study were included in any particular analysis.

5.2.5 Statistical Methods

I produced descriptive statistics at baseline for each LTC group. KM survival curves were charted and visually inspected to aid understanding of differences in survival across groups. Log-rank tests were used to test for significant differences across exposure cohorts. Trajectories of total and LTC-specific mean costs were charted for each LTC group, comparing cancer and non-cancer cohorts. Both yearly and phase-of-care trajectories were produced showing the mean costs in each year or phase relative to the diagnosis event. I chose to present cumulative costs in yearly post-diagnosis charts to better understand how the costs developed over time. Yearly charts were uncensored, meaning that deceased patients would contribute to the yearly means at rate zero. Phase-of-care costs were presented as specific to each phase rather than cumulative, as this highlighted where in the patient-timeline costs were being accrued. Phase-of-care costs only included patients who spent time in that phase, providing information on costs for living patients rather than the cohort in aggregate

(which incorporated the zero costs of dead patients). As times spent in each phase varied between individuals, monthly phase-of-care charts were also produced to show rates of resource use. The phases were equivalent to those defined in Section 3.2.5.

To estimate mean excess total costs and LTC-specific costs over the eight-year post-diagnosis period I used GLM regression models similar to those described in Section 4.2.4 for reasons described in that section. Crude and adjusted cost estimates using univariable and multivariable models respectively were produced. Excess costs in adjusted models took the interpretation of an exposure with confounders, where the exposure was the cancer diagnosis. The general model form was described in Section 4.2.4, with confounders age, sex, SIMD, network region, rurality, year of event, comorbidities as described in Section 5.2.4. As the modelling goal was to produce excess cost estimates adjusted for confounders, rather than analyse risk factors or produce predictive models, model parsimony was not a priority. Hence a non-parsimonious approach to variable selection was taken. Additionally, the distributions were similar in terms of age, sex, geographical region and SIMD due to residual matching of the original cohort dataset. Inspection of charts showed a non-linear relationship between age and costs, and the relationship varied between LTC groups. Fractional polynomial tests indicated a complex model would need to be specified to improve upon a five-category variable. As this would have required substantial processing time while differences in costs were <1%, I chose to use the simpler model using categories. This also provided information on how age contributed to the excess costs. While analysis of risk factors was not a goal of this study, the variation of results between LTC groups made this useful information and the benefits of its inclusion outweighed the small increase in accuracy.

Residual matching from the original dataset from which the cohorts were derived invalidated the assumption of independence of observations. Clustering was present in all LTC groups and particularly prominent (85%) in the no-comorbidity group, while others were <10%. Normal standard errors can be misleadingly large when clustering is present, leading to overly narrow confidence intervals, and robust standard errors may not compensate sufficiently [206]. Errors that adjust for clustered data may give biased results where the number of clusters is low [206]. No specific test for low numbers is recognised however a minimum of around 20–50 clusters is recommended [206]. As the lowest number of clusters in any LTC group was over 500, and preliminary tests gave no measured difference in the mean outcome but did show larger standard errors, I used clustered errors in all regression models

An issue in healthcare costing is the high mass at zero in outcome distributions. In chapters 3 and 4, the inclusion of prescriptions data and outpatient visits meant few zero-cost outcomes were measured. In this study, however, only SMR01 records were used, causing a higher proportion of zero costs, particularly in LTC-specific costs. GLMs are suitable for non-normal distributions but may provide lower precision than two-part models where very high numbers of zero counts are present [133]. An alternative method for data with high zero counts is the use of two-part models, in which the first part estimates the probability of a non-zero outcome, and the second estimates the magnitude of the outcome conditional on a non-zero outcome in part one. As a robustness check for the LTC-specific cost estimations, which had a higher proportion of zero counts than the total costs, I carried out two-part regressions using probit for the first part and gamma-log for the second part. Standard errors in two-part models that use retransformations must be obtained by bootstrap methods [207], however this was not the case here due to the use of GLMs in the second part. While two-part models could have been used as the primary estimation method, the reporting and interpretation of coefficients is less straightforward, hence GLM estimations remained the primary estimation method.

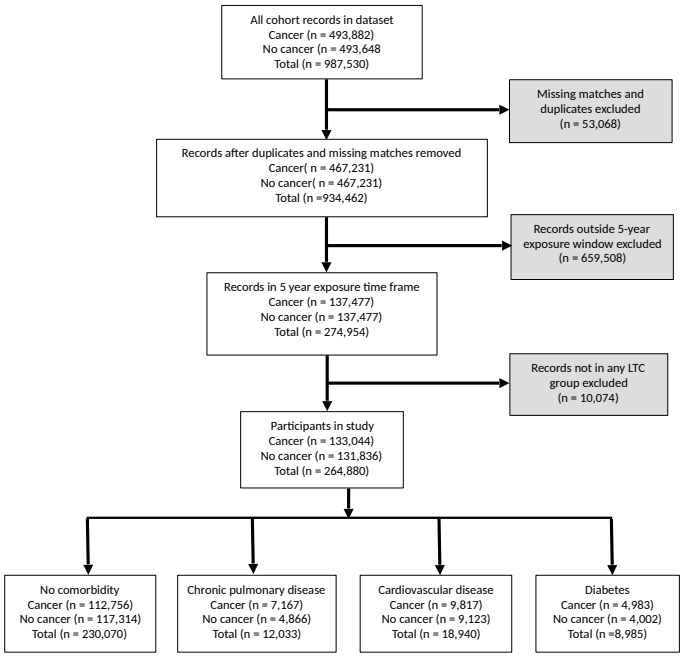
In all statistical tests 5% significance levels were used and I reported cost estimates with 95% confidence intervals. All analyses were carried out in Stata 16.0. Cost calculations were carried out using Stata's internal float precision but final costs were reported rounded to the nearest pound in MS Excel.

5.3 Results

5.3.1 Participants

After the removal of duplicates and implementation of inclusion criteria, a total of 264,880 patients, comprising 133,044 cancer patients and 131,836 non-cancer patients were included in the final analysis. Figure 5.1 shows how study numbers were derived. The no-comorbidity group (N=230,070) comprised 112,756 cancer and 117,314 non-cancer patients, the CVD group (N=18,940) 9,817 cancer patients and 9,123 non-cancer patients, the diabetes group (N=8,985) 4,983 cancer and 4,002 non-cancer patients, and the CPD group (N=12,033) included 7,167 cancer and 4,866 non-cancer patients. As patients could exist in multiple LTC groups (but not exposure groups) the LTC group numbers did not sum to the total number of participants.

Figure 5.1.: Flow chart of participant numbers



Notes: LTC groups overlap hence do not sum to study total.
Numbers prior to final selection represent database records rather than unique individuals.

5.3.2 Descriptive Data

Baseline demographics shown in Table 5.1 were similar for cancer and non-cancer cohorts across all LTC groups. The mean age for all patients was 67.4 (SD=13.8), largely driven by the no-comorbidity group with mean age 66.4 (14.0). The other LTC groups showed higher mean ages with 72.4 (11.1) for CPD, 75.6 (9.7) for CVD, and 73.0 (10.0) for diabetes. The no-comorbidity group and CPD groups had slightly higher proportions of female patients with 52.8% and 50.6% respectively, while in the CVD and diabetes groups, females comprised 39.1% and 43.2% of the respective groups. The West regional network, which incorporates the Glasgow conurbation, had the largest proportion of patients at 47.0% compared to 27.6% for South and East, and 25.4% for North, with a similar pattern of distribution in each LTC group. SIMD showed similar distributions across exposure cohorts, but marked differences across LTC groups, with higher proportions of LTC patients from deprived areas, particularly for CPD at 30.6%. The proportion of rural patients was similar in all LTC groups at 33–35% with no marked differences between exposure groups. The year of event showed a general trend of rising proportions with time, likely due to secular demographic changes, but the differences were minor (<2%).

Table 5.1.: Baseline characteristics of cancer and non-cancer cohorts in the LTC groups

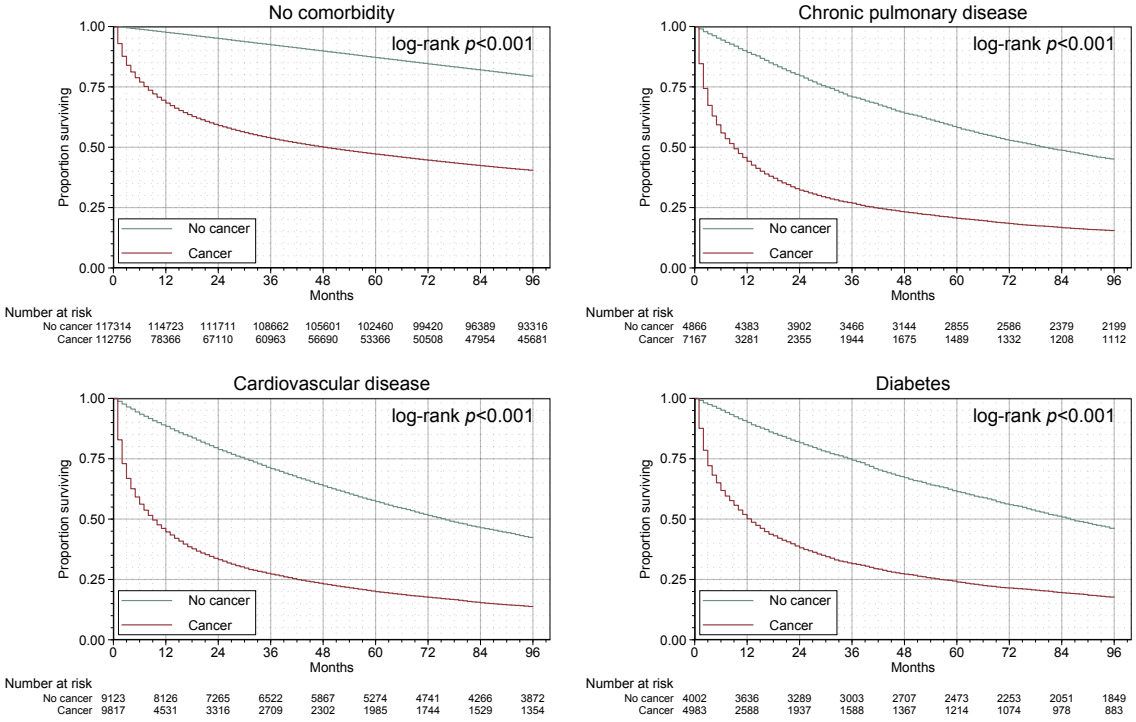
	No comorbidity		Chronic pulmonary disease		Cardiovascular disease		Diabetes		All patients		
	No cancer	Cancer	No cancer	Cancer	No cancer	Cancer	No cancer	Cancer	No cancer	Cancer	Total
	N=117,314	N=112,756	N=4,866	N=7,167	N=9,123	N=9,817	N=4,002	N=4,983	N=133,044	N=131,836	N=264,880
Sex											
Male	55,436 (47.3%)	53,173 (47.2%)	2,393 (49.2%)	3,548 (49.5%)	5,707 (62.6%)	5,833 (59.4%)	2,349 (58.7%)	2,758 (55.3%)	64,476 (48.5%)	63,648 (48.3%)	128,124 (48.4%)
Female	61,878 (52.7%)	59,583 (52.8%)	2,473 (50.8%)	3,619 (50.5%)	3,416 (37.4%)	3,984 (40.6%)	1,653 (41.3%)	2,225 (44.7%)	68,568 (51.5%)	68,188 (51.7%)	136,756 (51.6%)
Age in years (mean (sd))	66.4 (14.0)	66.3 (14.1)	72.5 (11.4)	72.4 (10.8)	75.9 (9.4)	75.4 (9.9)	73.4 (9.6)	72.7 (10.2)	67.4 (13.8)	67.4 (13.8)	67.4 (13.8)
Age categories											
< 50	14,012 (11.9%)	13,785 (12.2%)	222 (4.6%)	242 (3.4%)	70 (0.8%)	97 (1.0%)	70 (1.7%)	125 (2.5%)	14,354 (10.8%)	14,219 (10.8%)	28,573 (10.8%)
50-59	18,871 (16.1%)	18,161 (16.1%)	346 (7.1%)	563 (7.9%)	369 (4.0%)	548 (5.6%)	244 (6.1%)	380 (7.6%)	19,756 (14.8%)	19,507 (14.8%)	39,263 (14.8%)
60-69	31,386 (26.8%)	30,089 (26.7%)	10,55 (21.7%)	1,680 (23.4%)	1,746 (19.1%)	1,912 (19.5%)	930 (23.2%)	1,186 (23.8%)	34,681 (26.1%)	34,290 (26.0%)	68,971 (26.0%)
70-79	32,187 (27.4%)	30,689 (27.2%)	1,851 (38.0%)	2,786 (38.9%)	3,476 (38.1%)	3,673 (37.4%)	1,675 (41.9%)	2,010 (40.3%)	38,208 (28.7%)	37,921 (28.8%)	76,129 (28.7%)
>= 80	20,858 (17.8%)	20,032 (17.8%)	1,392 (28.6%)	1,896 (26.5%)	3,462 (37.9%)	3,587 (36.5%)	1,083 (27.1%)	1,282 (25.7%)	26,045 (19.6%)	25,899 (19.6%)	51,944 (19.6%)
Under 65	47,932 (40.9%)	46,423 (41.2%)	988 (20.3%)	1,461 (20.4%)	1,099 (12.0%)	1,367 (13.9%)	657 (16.4%)	990 (19.9%)	50,433 (37.9%)	49,854 (37.8%)	100,287 (37.9%)
SIMD quintile											
Least deprived	21,639 (18.4%)	21,046 (18.7%)	519 (10.7%)	697 (9.7%)	1,291 (14.2%)	1,439 (14.7%)	485 (12.1%)	592 (11.9%)	23,679 (17.8%)	23,478 (17.8%)	47,157 (17.8%)
2nd least deprived	22,362 (19.1%)	21,573 (19.1%)	735 (15.1%)	1,043 (14.6%)	1,560 (17.1%)	1,675 (17.1%)	628 (15.7%)	906 (18.2%)	24,978 (18.8%)	24,753 (18.8%)	49,731 (18.8%)
3rd least deprived	23,883 (20.4%)	22,906 (20.3%)	905 (18.6%)	1,376 (19.2%)	1,864 (20.4%)	1,976 (20.1%)	867 (21.7%)	1,040 (20.9%)	27,045 (20.3%)	26,739 (20.3%)	53,784 (20.3%)
2nd most deprived	25,106 (21.4%)	23,892 (21.2%)	1,231 (25.3%)	1,848 (25.8%)	2,147 (23.5%)	2,377 (24.2%)	944 (23.6%)	1,206 (24.2%)	28,863 (21.7%)	28,588 (21.7%)	57,451 (21.7%)
Most deprived	24,324 (20.7%)	23,339 (20.7%)	1,476 (30.3%)	2,203 (30.7%)	2,261 (24.8%)	2,350 (23.9%)	1,078 (26.9%)	1,239 (24.9%)	28,479 (21.4%)	28,278 (21.4%)	56,757 (21.4%)
Regional network											
South and east	32,290 (27.5%)	30,979 (27.5%)	1,455 (29.9%)	2,000 (27.9%)	2,602 (28.5%)	2,857 (29.1%)	1,122 (28.0%)	1,437 (28.8%)	36,766 (27.6%)	36,394 (27.6%)	73,160 (27.6%)
West	55,320 (47.2%)	53,669 (47.6%)	2,217 (45.6%)	3,237 (45.2%)	4,187 (45.9%)	4,210 (42.9%)	1,677 (41.9%)	2,005 (40.2%)	62,459 (46.9%)	61,945 (47.0%)	124,404 (47.0%)
North	29,704 (25.3%)	28,108 (24.9%)	1,194 (24.5%)	1,930 (26.9%)	2,334 (25.6%)	2,750 (28.0%)	1,203 (30.1%)	1,541 (30.9%)	33,819 (25.4%)	33,497 (25.4%)	67,316 (25.4%)
Rural location	40,217 (34.3%)	38,514 (34.2%)	1,568 (32.2%)	2,383 (33.2%)	3,032 (33.2%)	3,250 (33.1%)	1,496 (37.4%)	1,799 (36.1%)	45,503 (34.2%)	44,968 (34.1%)	90,471 (34.2%)
Year of diagnosis-event											
2006	22,682 (19.3%)	21,816 (19.3%)	880 (18.1%)	1,308 (18.3%)	1,812 (19.9%)	1,992 (20.3%)	700 (17.5%)	876 (17.6%)	25,676 (19.3%)	25,444 (19.3%)	51,120 (19.3%)
2007	23,133 (19.7%)	22,272 (19.8%)	945 (19.4%)	1,385 (19.3%)	1,851 (20.3%)	1,914 (19.5%)	745 (18.6%)	1,006 (20.2%)	26,239 (19.7%)	26,009 (19.7%)	52,248 (19.7%)
2008	23,652 (20.2%)	22,711 (20.1%)	978 (20.1%)	1,432 (20.0%)	1,811 (19.9%)	1,933 (19.7%)	828 (20.7%)	1,001 (20.1%)	26,796 (20.1%)	26,517 (20.1%)	53,313 (20.1%)
2009	23,985 (20.4%)	23,128 (20.5%)	1,023 (21.0%)	1,500 (20.9%)	1,896 (20.8%)	1,970 (20.1%)	892 (22.3%)	1,020 (20.5%)	27,291 (20.5%)	27,039 (20.5%)	54,330 (20.5%)
2010	23,862 (20.3%)	22,829 (20.2%)	1,040 (21.4%)	1,542 (21.5%)	1,753 (19.2%)	2,008 (20.5%)	837 (20.9%)	1,080 (21.7%)	27,042 (20.3%)	26,827 (20.3%)	53,869 (20.3%)
Survival months* (mean (sd))	86.3 (22.8)	51.8 (41.1)	64.5 (34.6)	27.6 (34.7)	63.6 (34.6)	27.4 (34.1)	66.7 (33.8)	31.6 (36.0)	83.8 (25.3)	48.5 (41.1)	66.3 (38.4)
Previous SMR01 episodes (mean (sd))	2.9 (5.2)	3.7 (5.8)	7.6 (8.8)	8.3 (9.1)	7.3 (7.8)	7.8 (7.5)	7.8 (9.0)	8.6 (12.2)	3.4 (5.7)	4.3 (6.5)	3.8 (6.2)
Charlson Comorbidity Score (mean (sd))	NA	NA	1.4 (0.8)	1.4 (0.7)	1.5 (0.8)	1.6 (0.9)	1.7 (0.9)	1.7 (0.9)	0.2 (0.5)	0.2 (0.6)	0.2 (0.5)
Comorbidities (mean (sd))	NA	NA	1.4 (0.7)	1.4 (0.6)	1.5 (0.7)	1.5 (0.7)	1.6 (0.7)	1.5 (0.7)	0.2 (0.5)	0.2 (0.5)	0.2 (0.5)
Acute myocardial infarcation	NA	NA	325 (6.7%)	438 (6.1%)	3,341 (36.6%)	3,305 (33.7%)	401 (10.0%)	403 (8.1%)	3,341 (2.5%)	3,305 (2.5%)	6,646 (2.5%)
Congestive heart failure	NA	NA	308 (6.3%)	371 (5.2%)	2,047 (22.4%)	2,163 (22.0%)	261 (6.5%)	279 (5.6%)	2,047 (1.5%)	2,163 (1.6%)	4,210 (1.6%)
Peripheral vascular disease	NA	NA	156 (3.2%)	310 (4.3%)	1,610 (17.6%)	2,235 (22.8%)	242 (6.0%)	280 (5.6%)	1,610 (1.2%)	2,235 (1.7%)	3,845 (1.5%)
Cerebral vascular disease	NA	NA	216 (4.4%)	300 (4.2%)	3,221 (35.3%)	3,224 (32.8%)	334 (8.3%)	320 (6.4%)	3,221 (2.4%)	3,224 (2.4%)	6,445 (2.4%)
Dementia	NA	NA	115 (2.4%)	147 (2.1%)	339 (3.7%)	262 (2.7%)	145 (3.6%)	113 (2.3%)	530 (0.4%)	461 (0.3%)	991 (0.4%)
Chronic pulmonary disease	NA	NA	4,866 (100.0%)	7,167 (100.0%)	863 (9.5%)	1,256 (12.8%)	440 (11.0%)	635 (12.7%)	4,866 (3.7%)	7167 (5.4%)	12,033 (4.5%)
Rheumatoid disease	NA	NA	96 (2.0%)	122 (1.7%)	112 (1.2%)	118 (1.2%)	46 (1.1%)	63 (1.3%)	225 (0.2%)	272 (0.2%)	497 (0.2%)
Peptic ulcer	NA	NA	46 (0.9%)	59 (0.8%)	61 (0.7%)	87 (0.9%)	30 (0.7%)	51 (1.0%)	125 (0.1%)	171 (0.1%)	296 (0.1%)
Diabetes	NA	NA	440 (9.0%)	635 (8.9%)	1,067 (11.7%)	1,119 (11.4%)	4,002 (100.0%)	4,983 (100.0%)	4,002 (3.0%)	4,983 (3.8%)	8,985 (3.4%)
Renal disease - moderate or severe	NA	NA	144 (3.0%)	231 (3.2%)	415 (4.5%)	553 (5.6%)	241 (6.0%)	338 (6.8%)	654 (0.5%)	910 (0.7%)	1,564 (0.6%)
Other comorbidity	NA	NA	47 (1.0%)	124 (1.7%)	260 (2.8%)	290 (3.0%)	57 (1.4%)	155 (3.1%)	320 (0.2%)	500 (0.4%)	820 (0.3%)

Notes: SIMD = Scottish Index of Multiple Deprivation, sd = standard deviation. Characteristics recorded at time of diagnosis event or prior to in the case of comorbidities. LTC groups overlap hence numbers may not sum to the study total.

5.3.3 Outcomes Data

Survival was considerably poorer for cancer patients as expected, and also for patients with LTCs compared to patients with no comorbidities. Figure 5.2 shows higher eight-year survival for non-cancer patients in all LTC groups. The curves for patients with LTCs were similar to each other, with very low survival for cancer patients at eight years and survival for non-cancer patients under 50%, which was similar to cancer patients in the no-comorbidity group.

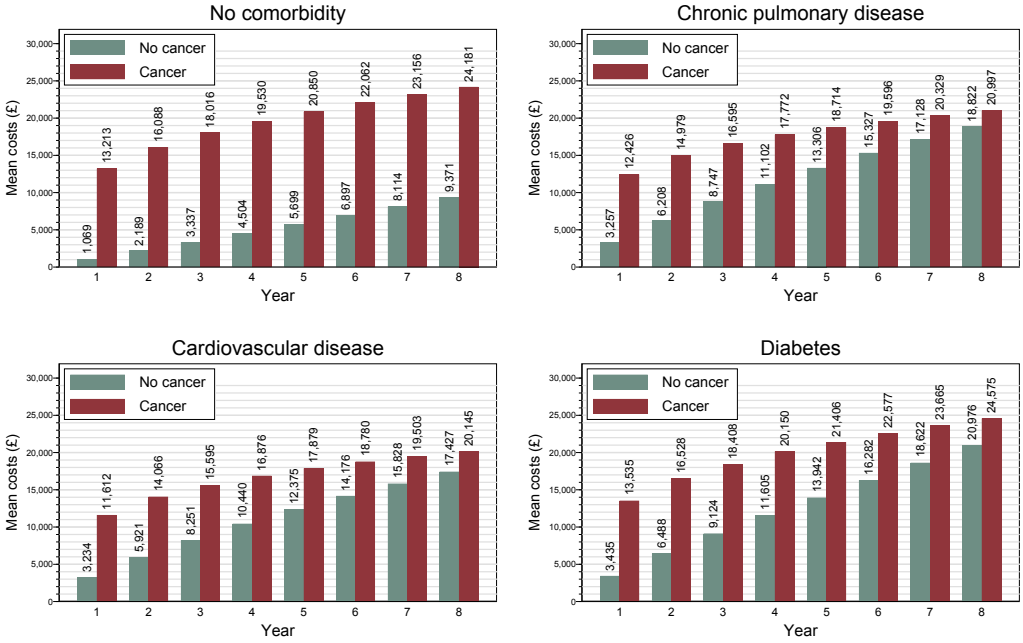
Figure 5.2.: Eight-year Kaplan Meier survival trajectories of cancer and non-cancer cohorts across LTC groups



Notes: Survival times were derived from SMR06 and NRS Death records rounded to the nearest month.

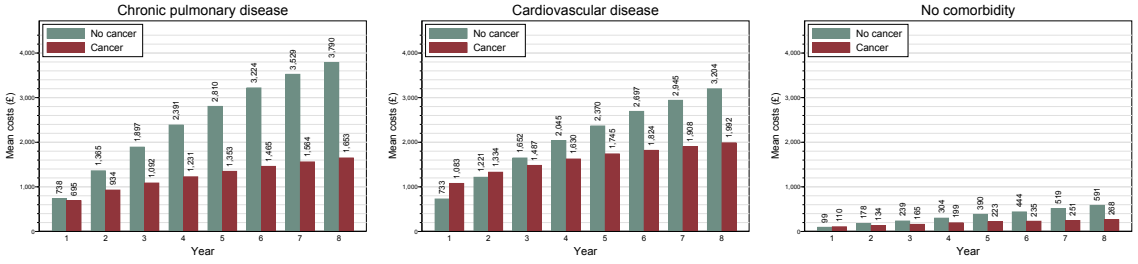
Trajectories of cumulative yearly costs in Figure 5.3 show greater costs for the cancer cohort at all periods for all LTC groups. However, there was considerable variation across LTC groups. Cancer was associated with substantially higher total costs in the no-comorbidity group in all periods. Cancer was also associated with higher total costs in the groups with LTCs, but the association was considerably lower than observed in the no-comorbidity group. LTC-specific costs in Figure 5.4 were lower in the cancer cohorts over all years and the differential between cancer and non-cancer cohorts increased over time. This figure also shows that diabetes-specific costs were substantially lower than other LTC-specific costs for cancer and non-cancer patients.

Figure 5.3.: Cumulative eight-year inpatient costs for cancer and non-cancer cohorts across LTC groups



Notes: All costs derived from SMR01 (inpatient and daycase) records. Costs are undiscounted in pounds sterling at 2018 levels. Years are relative to the diagnosis-event. All patients represented in all years. Deceased patients incurred costs at zero.

Figure 5.4.: Cumulative eight-year LTC-specific inpatient costs for cancer and non-cancer cohorts across LTC groups

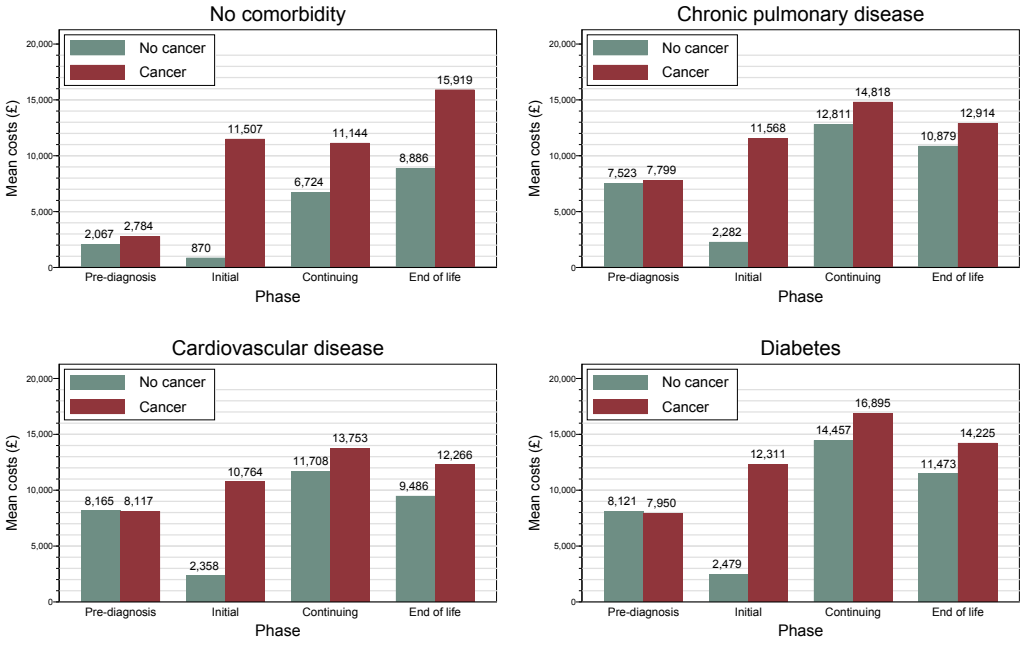


Notes: The no-comorbidity group is not presented as the figure shows LTC-specific costs only. All costs derived from SMR01 (inpatient and daycase) records where LTC was listed as the main condition. Costs are undiscounted in pounds sterling at 2018 levels. Years are relative to the diagnosis-event. All patients represented in all years. Deceased patients incurred costs at zero.

Figure 5.5 shows costs by phase-of-care relative to the diagnosis (or pseudo-diagnosis event for the non-cancer cohort). During the initial phase the expected post-diagnosis spikes were observed in all cancer groups, with smaller though still notable differences in the continuing and end-of-life periods. Phase-of-care costs provide a different perspective to annual costs as they include only patients who entered the phase, and the times in each phase varied by patient. Monthly phase-of-care costs in Figure 5.6

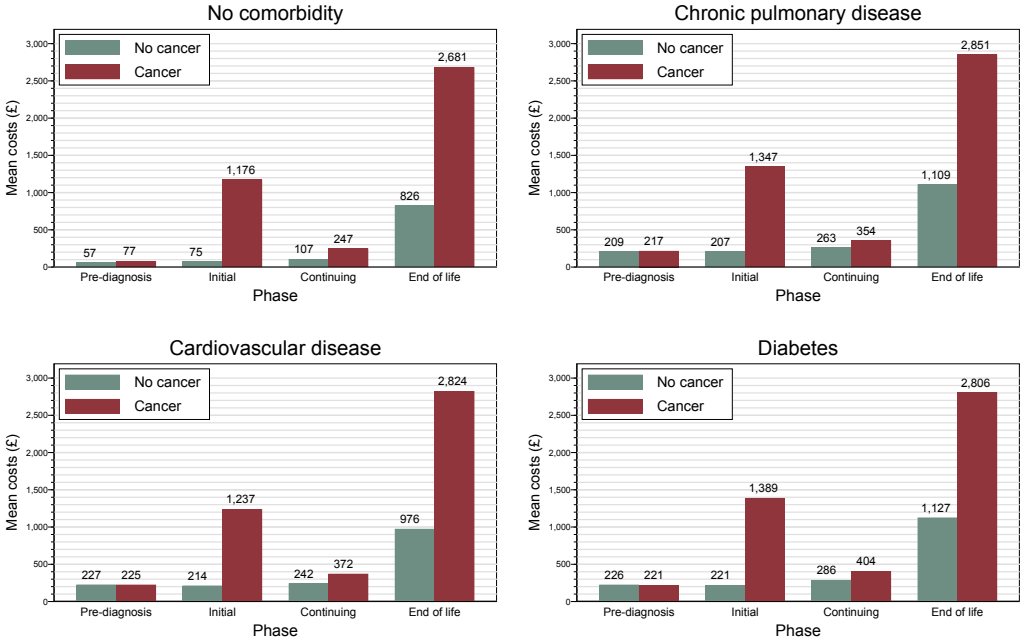
show the rates of cost accrual over each phase. The rates for all LTC groups were higher than those of the no-comorbidity group in all post-diagnosis phases. This was also the case for cancer patients compared to non-cancer patients. Rates of pre-diagnosis costs were similar for cancer and non-cancer patients with LTCs, but higher for cancer patients in the no-comorbidity group, though of relatively low magnitudes for both. Figure 5.7 shows LTC-specific costs by phase-of-care. As in the yearly cumulative total costs, higher costs were observed in the non-cancer cohort for all LTC groups. Monthly phase-of-care LTC-specific costs are shown in Figure 5.8. Only CVD in the initial and end-of-life phases and diabetes in the initial phase showed substantially higher costs for cancer patients.

Figure 5.5.: Phase-of-care total inpatient costs for cancer and non-cancer cohorts across LTC groups



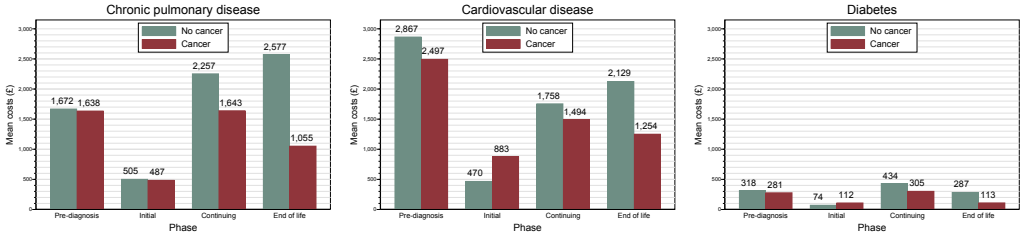
Notes: All costs derived from SMR01 (inpatient and daycase) records. Costs are undiscounted in pounds sterling at 2018 levels. Phases are relative to the diagnosis-event. Only patients who entered a phase were included.

Figure 5.6.: Monthly phase-of-care inpatient costs for cancer and non-cancer cohorts across LTC groups



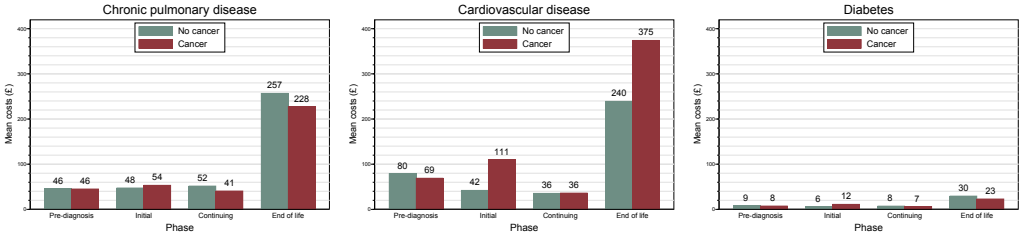
Notes: All costs derived from SMR01 (inpatient and daycase) records. Costs are undiscounted in pounds sterling at 2018 levels. Phases are relative to the diagnosis-event. Only patients who entered a phase were included.

Figure 5.7.: Phase-of-care LTC-specific inpatient costs for cancer and non-cancer cohorts across LTC groups



Notes: The no-comorbidity group is not presented as the figure shows LTC-specific costs only. All costs derived from SMR01 (inpatient and daycase) records where the LTC was listed as main condition. Costs are undiscounted in pounds sterling at 2018 levels. Phases are relative to the diagnosis-event. Only patients who entered a phase were included.

Figure 5.8.: Monthly phase-of-care LTC-specific inpatient costs for cancer and non-cancer cohorts across LTC groups



Notes: The no-comorbidity group is not presented as the figure shows LTC-specific costs only. All costs derived from SMR01 (inpatient and daycase) records where the LTC was listed as main condition. Costs are undiscounted in pounds sterling at 2018 levels. Phases are relative to the diagnosis-event. Only patients who entered a phase were included.

5.3.4 Results of Regression Analyses

Table 5.2 shows that cancer was significantly associated with increased costs at the 5% significance level in all groups. The largest association was seen in the no-comorbidity group with an adjusted cost ratio of 3.016 (95%CI 2.975 to 3.058), suggesting cancer raised costs by over 200%. This equated to an adjusted cost of £17,555 (95%CI £17,322 to £17,788) undiscounted at 2017/18 price levels, as shown in Table 5.4. These costs were substantially greater than those in patients with LTCs. The adjusted cost ratios were CPD 1.069 (1.023 to 1.118), CVD 1.125 (1.086 to 1.165), diabetes 1.128 (1.075 to 1.185). These translated into undiscounted costs of £1,341 (£459 to £2,223) for CPD, £2,208 (£1,552 to £2,863) for CVD and £2,751 (£1,653 to £3,849) for diabetes, at 2017/18 price levels after adjustment for confounders. The association was reversed for LTC-specific costs, with the cost ratios in Table 5.3 below 1 for all LTCs ($p<0.001$), indicating a negative association between cancer and LTC-specific excess costs in each group. The relative reduction was greatest in diabetes with an adjusted CR of 0.361 (95% CI 0.277 to 0.471), and lowest in CVD with adjusted CR 0.630 (0.587 to 0.677), and CPD with an adjusted CR of 0.450 (0.404 to 0.502). However, in monetary terms, diabetes had the lowest costs. The coefficients translated into monetary costs of -£2,073 (95%CI -£2,367 to -£1,780) for CPD, -£1,176 (-£1,359 to -£922) for CVD, and -£452 (-£591 to -£314) for diabetes.

Table 5.2.: Crude and adjusted GLM regressions with eight-year costs as dependent variable across LTC groups

	No comorbidity								Chronic pulmonary disease							
	Crude				Adjusted				Crude				Adjusted			
Variable	CR	*p*	95% CI		CR	*p*	95% CI		CR	*p*	95% CI (exp)		CR	*p*	95% CI	
Cancer	2.580	<0.001	2.549	2.613	3.016	<0.001	2.975	3.058	1.116	<0.001	1.066	1.167	1.069	0.003	1.023	1.118
Female	--	--	--	--	0.922	<0.001	0.910	0.933	--	--	--	--	1.023	0.291	0.981	1.067
Age																
< 50	--	--	--	--	Ref	--	--	--	--	--	--	--	Ref	--	--	--
50-59	--	--	--	--	1.069	<0.001	1.041	1.097	--	--	--	--	0.989	0.898	0.840	1.165
60-69	--	--	--	--	1.227	<0.001	1.199	1.257	--	--	--	--	0.867	0.065	0.744	1.009
70-79	--	--	--	--	1.605	<0.001	1.567	1.643	--	--	--	--	0.838	0.021	0.721	0.974
>= 80	--	--	--	--	1.731	<0.001	1.688	1.776	--	--	--	--	0.679	<0.001	0.584	0.791
Year of event																
2006	--	--	--	--	Ref	--	--	--	--	--	--	--	Ref	--	--	--
2007	--	--	--	--	1.021	0.040	1.001	1.041	--	--	--	--	1.033	0.349	0.966	1.104
2008	--	--	--	--	1.036	<0.001	1.017	1.056	--	--	--	--	1.035	0.314	0.968	1.107
2009	--	--	--	--	1.040	<0.001	1.020	1.060	--	--	--	--	1.046	0.180	0.979	1.118
2010	--	--	--	--	1.052	<0.001	1.031	1.072	--	--	--	--	1.062	0.074	0.994	1.134
Region network																
West					Ref				--	--	--	--	Ref	--	--	--
South and east					0.940	<0.001	0.925	0.955	--	--	--	--	1.015	0.593	0.962	1.070
North					0.876	<0.001	0.862	0.889	--	--	--	--	0.977	0.407	0.925	1.032
Rural					0.964	<0.001	0.951	0.978					1.013	0.593	0.965	1.064
SIMD																
Least deprived	--	--	--	--	Ref	--	--	--	--	--	--	--	Ref	--	--	--
2nd least deprived	--	--	--	--	1.018	0.084	0.998	1.039	--	--	--	--	0.928	0.088	0.852	1.011
3rd least deprived	--	--	--	--	1.084	<0.001	1.063	1.106	--	--	--	--	0.959	0.324	0.883	1.042
2nd most deprived	--	--	--	--	1.156	<0.001	1.133	1.179	--	--	--	--	0.972	0.474	0.901	1.050
Most deprived	--	--	--	--	1.202	<0.001	1.178	1.227	--	--	--	--	0.995	0.897	0.921	1.075
AMI	--	--	--	--	--	--	--	--	--	--	--	--	1.009	0.835	0.928	1.097
CHF	--	--	--	--	--	--	--	--	--	--	--	--	0.872	0.004	0.795	0.957
PVD	--	--	--	--	--	--	--	--	--	--	--	--	1.024	0.665	0.919	1.141
CEVD	--	--	--	--	--	--	--	--	--	--	--	--	0.911	0.071	0.824	1.008
Dementia	--	--	--	--	--	--	--	--	--	--	--	--	0.581	<0.001	0.511	0.662
COPD	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
Rheum	--	--	--	--	--	--	--	--	--	--	--	--	1.158	0.134	0.956	1.404
PUD	--	--	--	--	--	--	--	--	--	--	--	--	0.790	0.010	0.659	0.946
Diabetes	--	--	--	--	--	--	--	--	--	--	--	--	1.281	<0.001	1.192	1.377
Renal	--	--	--	--	--	--	--	--	--	--	--	--	1.097	0.120	0.976	1.233
Other comorbidity	--	--	--	--	--	--	--	--	--	--	--	--	0.967	0.679	0.826	1.132
Constant	9371	<0.001	9275	9467	6248	<0.001	6055	6446	18822	<0.001	18156	19512	22803	<0.001	19181	27109

	Cardiovascular disease								Diabetes							
	Crude				Adjusted				Crude				Adjusted			
Variable	CR	*p*	95% CI		CR	*p*	95% CI		CR	*p*	95% CI		CR	*p*	95% CI	
Cancer	1.156	<0.001	1.115	1.198	1.125	<0.001	1.086	1.165	1.172	<0.001	1.111	1.235	1.128	<0.001	1.075	1.185
Female	--	--	--	--	0.978	0.214	0.944	1.013	--	--	--	--	0.996	0.880	0.949	1.046
Age																
< 50	--	--	--	--	Ref	--	--	--	--	--	--	--	Ref	--	--	--
50-59	--	--	--	--	0.953	0.677	0.758	1.197	--	--	--	--	0.796	0.158	0.580	1.093
60-69	--	--	--	--	0.855	0.150	0.690	1.058	--	--	--	--	0.722	0.037	0.532	0.981
70-79	--	--	--	--	0.827	0.077	0.669	1.021	--	--	--	--	0.718	0.034	0.529	0.975
>= 80	--	--	--	--	0.698	0.001	0.565	0.863	--	--	--	--	0.550	<0.001	0.405	0.746
Year of event																
2006	--	--	--	--	Ref	--	--	--					Ref	--	--	--
2007	--	--	--	--	1.009	0.747	0.956	1.064	--	--	--	--	1.032	0.413	0.956	1.114
2008	--	--	--	--	1.030	0.266	0.978	1.085	--	--	--	--	1.043	0.268	0.968	1.124
2009	--	--	--	--	1.038	0.175	0.983	1.096	--	--	--	--	1.048	0.198	0.976	1.127
2010	--	--	--	--	1.121	<0.001	1.062	1.183	--	--	--	--	1.108	0.007	1.029	1.194
Region network																
West	--	--	--	--	Ref	--	--	--	--	--	--	--	Ref	--	--	--
South and east	--	--	--	--	1.043	0.054	0.999	1.089	--	--	--	--	1.007	0.823	0.949	1.068
North	--	--	--	--	0.945	0.014	0.904	0.989	--	--	--	--	0.928	0.013	0.875	0.985
Rural					0.948	0.010	0.910	0.987	--	--	--	--	0.966	0.193	0.918	1.017
SIMD																
Least deprived	--	--	--	--	Ref	--	--	--	--	--	--	--	Ref	--	--	--
2nd least deprived	--	--	--	--	1.004	0.899	0.943	1.069	--	--	--	--	0.952	0.246	0.875	1.035
3rd least deprived	--	--	--	--	1.019	0.535	0.960	1.081	--	--	--	--	1.016	0.723	0.932	1.107
2nd most deprived	--	--	--	--	1.032	0.297	0.973	1.093	--	--	--	--	1.060	0.155	0.978	1.150
Most deprived	--	--	--	--	1.041	0.177	0.982	1.104	--	--	--	--	1.031	0.482	0.947	1.121
AMI	--	--	--	--	--	--	--	--	--	--	--	--	1.063	0.185	0.971	1.164
CHF	--	--	--	--	--	--	--	--	--	--	--	--	0.896	0.029	0.812	0.989
PVD	--	--	--	--	--	--	--	--	--	--	--	--	1.098	0.049	1.000	1.204
CEVD	--	--	--	--	--	--	--	--	--	--	--	--	0.879	0.007	0.801	0.966
Dementia	--	--	--	--	0.546	<0.001	0.491	0.607	--	--	--	--	0.533	<0.001	0.454	0.625
COPD	--	--	--	--	0.987	0.632	0.936	1.041	--	--	--	--	1.102	0.009	1.025	1.185
Rheum	--	--	--	--	1.095	0.195	0.955	1.256	--	--	--	--	0.953	0.565	0.810	1.122
PUD	--	--	--	--	0.914	0.321	0.765	1.092	--	--	--	--	1.131	0.247	0.918	1.393
Diabetes	--	--	--	--	1.174	<0.001	1.115	1.236	--	--	--	--	--	--	--	--
Renal	--	--	--	--	1.126	0.018	1.020	1.243	--	--	--	--	1.225	0.002	1.079	1.391
Other comorbidity	--	--	--	--	0.886	0.014	0.804	0.976	--	--	--	--	0.881	0.149	0.741	1.047
Constant	17427	<0.001	16963	17904	21410	<0.001	17142	26741	20976	<0.001	20157	21829	29995	<0.001	22098	40713

Abbreviations: GLM=Generalised Linear Model, CR=cost ratio, SIMD=Scottish Index of Multiple Deprivation, AMI=Acute myocardial infarcation, CHF=Congestive heart failure, PVD=Peripheral vascular disease, CEVD=Cerebral vascular disease, PUD=Peptic ulcer, COPD=Chronic pulmonary disease, Rheumd=Rheumatoid disease.
Notes: Coefficients are represented as exponents.

Table 5.3.: Crude and adjusted GLM regressions with eight-year LTC-specific costs as dependent variable across LTC groups

	Chronic pulmonary disease								Cardiovascular disease							
	Crude				Adjusted				Crude				Adjusted			
	CR	*p*	95% CI		CR	*p*	95% CI		CR	*p*	95% CI		CR	*p*	95% CI	
Cancer	0.436	<0.001	0.391	0.487	0.446	<0.001	0.401	0.497	0.622	<0.001	0.579	0.668	0.630	<0.001	0.587	0.677
Female	--	--	--	--	0.997	0.964	0.891	1.117	--	--	--	--	0.964	0.334	0.896	1.038
Age																
< 50	--	--	--	--	Ref	--	--	--	--	--	--	--	Ref	--	--	--
50-59	--	--	--	--	1.797	0.012	1.137	2.841	--	--	--	--	0.580	0.027	0.357	0.941
60-69	--	--	--	--	1.902	0.001	1.284	2.816	--	--	--	--	0.797	0.331	0.505	1.259
70-79	--	--	--	--	2.039	<0.001	1.379	3.016	--	--	--	--	0.850	0.481	0.541	1.336
>= 80	--	--	--	--	1.449	0.064	0.978	2.148	--	--	--	--	0.809	0.357	0.515	1.271
Year of event																
2006	--	--	--	--	Ref	--	--	--	--	--	--	--	Ref	--	--	--
2007	--	--	--	--	1.133	0.156	0.954	1.347	--	--	--	--	0.945	0.320	0.846	1.056
2008	--	--	--	--	1.067	0.465	0.896	1.271	--	--	--	--	1.027	0.650	0.916	1.152
2009	--	--	--	--	1.051	0.572	0.884	1.249	--	--	--	--	1.005	0.929	0.900	1.122
2010	--	--	--	--	1.069	0.418	0.909	1.257	--	--	--	--	1.134	0.029	1.013	1.269
	--	--	--	--					--	--	--	--				
Region network	--	--	--	--					--	--	--	--				
West	--	--	--	--	Ref	--	--	--	--	--	--	--	Ref	--	--	--
South and east	--	--	--	--	0.899	0.114	0.788	1.026	--	--	--	--	1.132	0.005	1.039	1.234
North	--	--	--	--	0.911	0.224	0.785	1.058	--	--	--	--	1.108	0.038	1.006	1.222
	--	--	--	--					--	--	--	--				
Rural	--	--	--	--	0.825	0.003	0.727	0.935	--	--	--	--	0.923	0.063	0.849	1.004
SIMD																
Least deprived	--	--	--	--	Ref	--	--	--	--	--	--	--	Ref	--	--	--
2nd least deprived	--	--	--	--	1.138	0.308	0.887	1.460	--	--	--	--	1.020	0.770	0.895	1.162
3rd least deprived	--	--	--	--	1.184	0.156	0.938	1.494	--	--	--	--	1.081	0.218	0.955	1.225
2nd most deprived	--	--	--	--	1.239	0.054	0.996	1.542	--	--	--	--	1.062	0.318	0.943	1.196
Most deprived	--	--	--	--	1.422	0.002	1.139	1.774	--	--	--	--	1.054	0.403	0.932	1.192
AMI	--	--	--	--	1.019	0.864	0.822	1.263	--	--	--	--	--	--	--	--
CHF	--	--	--	--	0.802	0.036	0.653	0.986	--	--	--	--	--	--	--	--
PVD	--	--	--	--	1.282	0.184	0.889	1.850	--	--	--	--	--	--	--	--
CEVD	--	--	--	--	0.727	0.038	0.538	0.982	--	--	--	--	--	--	--	--
Dementia	--	--	--	--	0.674	0.023	0.479	0.948	--	--	--	--	0.383	<0.001	0.304	0.482
COPD	--	--	--	--	--	--	--	--	--	--	--	--	0.749	<0.001	0.662	0.848
Rheum	--	--	--	--	0.909	0.573	0.653	1.266	--	--	--	--	0.990	0.956	0.695	1.410
PUD	--	--	--	--	0.419	0.001	0.253	0.696	--	--	--	--	0.747	0.275	0.442	1.262
Diabetes	--	--	--	--	0.873	0.274	0.684	1.114	--	--	--	--	1.266	<0.001	1.144	1.402
Renal	--	--	--	--	0.873	0.348	0.658	1.159	--	--	--	--	1.366	<0.001	1.179	1.583
Other comorbidity	--	--	--	--	0.665	0.043	0.448	0.988	--	--	--	--	0.699	0.005	0.543	0.900
Constant	3,790	<0.001	3,533	4,065	1,837	<0.001	1,160	2,911	3,204	<0.001	3,060	3,354	3,601	<0.001	2,249	5,764

	Diabetes							
	Crude				Adjusted			
	CR	*p*	95% CI		CR	*p*	95% CI	
Cancer	0.453	<0.001	0.344	0.597	0.361	<0.001	0.277	0.471
Female	--	--	--	--	1.286	0.053	0.997	1.658
Age								
< 50	--	--	--	--	Ref	--	--	--
50-59	--	--	--	--	0.511	0.070	0.247	1.056
60-69	--	--	--	--	0.188	<0.001	0.100	0.354
70-79	--	--	--	--	0.217	<0.001	0.120	0.394
>= 80	--	--	--	--	0.151	<0.001	0.083	0.275
Year of event								
2006	--	--	--	--	Ref	--	--	--
2007	--	--	--	--	0.976	0.895	0.680	1.402
2008	--	--	--	--	1.382	0.125	0.914	2.090
2009	--	--	--	--	0.795	0.235	0.543	1.162
2010	--	--	--	--	0.935	0.758	0.607	1.439
	--	--	--	--				
Region network	--	--	--	--				
West	--	--	--	--	Ref	--	--	--
South and east	--	--	--	--	0.923	0.631	0.664	1.282
North	--	--	--	--	1.132	0.446	0.823	1.558
	--	--	--	--				
Rural	--	--	--	--	1.023	0.877	0.765	1.368
SIMD								
Least deprived	--	--	--	--	Ref	--	--	--
2nd least deprived	--	--	--	--	1.645	0.086	0.931	2.906
3rd least deprived	--	--	--	--	0.889	0.676	0.512	1.543
2nd most deprived	--	--	--	--	1.234	0.433	0.730	2.086
Most deprived	--	--	--	--	1.067	0.809	0.631	1.804
AMI	--	--	--	--	1.129	0.604	0.714	1.787
CHF	--	--	--	--	0.994	0.980	0.611	1.617
PVD	--	--	--	--	3.429	<0.001	2.204	5.336
CEVD	--	--	--	--	0.824	0.371	0.540	1.258
Dementia	--	--	--	--	0.594	0.145	0.295	1.197
COPD	--	--	--	--	0.649	0.031	0.438	0.961
Rheum	--	--	--	--	0.024	<0.001	0.005	0.125
PUD	--	--	--	--	0.098	0.025	0.013	0.752
Diabetes	--	--	--	--	--	--	--	--
Renal	--	--	--	--	0.777	0.274	0.495	1.221
Other comorbidity	--	--	--	--	0.040	<0.001	0.011	0.148
Constant	591	<0.001	501	696	2,235	<0.001	1,021	4,892

Abbreviations: GLM=Generalised Linear Model, CR=cost ratio, SIMD=Scottish Index of Multiple Deprivation, AMI=Acute myocardial infarcation, CHF=Congestive heart failure, PVD=Peripheral vascular disease, CEVD=Cerebral vascular disease, PUD=Peptic ulcer, COPD=Chronic pulmonary disease, Rheumd=Rheumatoid disease.
Notes: Coefficients are represented as exponents.

Table 5.4.: Eight-year cancer cost estimates from average marginal effects of crude and adjusted GLM and two-part model regressions across LTC groups

	Crude					Adjusted				
LTC group	**dydx (£)**	**Std. error**	***p***	**95%CI (lower/upper)**		**dydx (£)**	**Std. error**	***p***	**95%CI (lower/upper)**	
TOTAL COSTS										
No comorbidity	14,810	98.049	<0.001	14,618	15,002	17,555	118.853	<0.001	17,322	17,788
CPD	2,175	452.181	<0.001	1,289	3,062	1,341	449.845	0.003	459	2,223
CVD	2,718	343.536	<0.001	2,045	3,391	2,208	334.486	<0.001	1,552	2,863
Diabetes	3,599	609.044	<0.001	2,405	4,792	2,751	560.176	<0.001	1,653	3,849
LTC-SPECIFIC GLM										
CPD	-2,137	153.548	<0.001	-2,438	-1,836	-2,073	149.787	<0.001	-2,367	-1,780
CVD	-1,212	93.397	<0.001	-1,395	-1,028	-1,176	93.516	<0.001	-1,359	-992
Diabetes	-323	58.019	<0.001	-437	-209	-452	70.788	<0.001	-591	-314
LTC-SPECIFIC 2PM										
CPD	-2,137	153.526	<0.001	-2,438	-1,836	-2,164	149.052	<0.001	-2,457	-1,872
CVD	-1,212	93.319	<0.001	-1,394	-1,029	-1,212	92.160	<0.001	-1,393	-1,032
Diabetes	-323	58.046	<0.001	-437	-209	-344	52.115	<0.001	-446	-242

Abbreviations: GLM=Generalised Linear Model, LTC=long-term condition, 2PM=two-part model, CI=confidence interval, CVD=Cardiovascular disease, CPD=Chronic pulmonary disease, dydx=marginal effect.
Notes: Costs derived from marginal effects on cancer variable with other variables at means. All costs derived from SMR01 (inpatient and daycase) records. Costs are undiscounted in pounds sterling at 2018 levels.

5.4 Discussion

5.4.1 Key Results

In this analysis I charted survival trajectories and inpatient cost trajectories of cancer and non-cancer patients with LTCs, and measured eight-year total excess costs and eight-year LTC-specific excess costs for cancer patients. Survival trajectories showed lower survival for patients with LTCs compared to patients without comorbidities, and for cancer patients compared to non-cancer patients. Trajectories of total costs showed higher healthcare use in patients with LTCs than without, and in cancer patients than non-cancer patients. Cost accumulation rates were higher for patients with LTCs regardless of cancer status, and for cancer patients across all LTC groups in all phases of care and years after diagnosis. Diabetes-specific costs were substantially lower than costs specific to CPD and CVD in all phases of care and years after diagnosis. After adjustment for confounders, eight-year excess costs of cancer were positive for all groups and highest in patients without comorbidities, while eight-year LTC-specific excess costs were negative for all LTCs. Excess costs in the no-comorbidity group were highest both relative in terms of a ratio of baseline costs and in terms of absolute monetary units.

5.4.2 Interpretation

Patterns of Costs Over Time

Cost trajectories showed similar patterns across LTCs, both in terms of trends over time and in terms of differences between cancer and non-cancer patients, while being markedly different from those in the no-comorbidity group. Survival trajectories were also similar across LTCs but much lower survival was observed in the no-comorbidity group for both cancer and non-cancer patients. However the relationship between survival and costs was non-linear, with non-cancer individuals in the no-comorbidity group having the highest survival and the lowest eight-year costs. Common features of all LTC cohorts were lower survival and higher total costs in cancer patients, with rates of cost accumulation and total accumulated costs higher in all post-diagnosis phases of care. Pre-diagnosis costs were similar for cancer and non-cancer groups, suggesting that cancer increased total costs in all phases. The lack of differences in pre-diagnosis costs also provides evidence of similarity across cancer and non-cancer cohorts.

Patterns of Costs Across Cohorts

Cumulative costs increased more rapidly for non-cancer patients with LTCs than for cancer patients with LTCs, making it plausible that excess costs could have turned negative over longer time frames than the eight years of follow-up. However, for patients without comorbidities, excess costs increased over time despite higher mortality in cancer patients. While the absolute rate of cost accumulation was lower in all phases for this group than for cancer patients with LTCs, the lower healthcare use of patients without cancer or comorbidities increased the relative disparity, suggesting the impact of cancer was strongest for this group and did not diminish over time. Cumulative costs for cancer patients in the no comorbidity group were higher at all time points than the CVD and CPD groups, but lower than the diabetes group. Lower costs in the care phases for the no-comorbidity group suggested higher survival as an explanation, yet this did not hold for diabetes, where the rates of cost accumulation were higher than other groups except in the end-of-life period. This may have been a result of higher levels of comorbidity in the diabetes group, as the average Charlson score was slightly higher than in the CPD and CVD groups. Variation in the stage of detection and subsequent treatment could explain higher costs for cancer patients, however diabetes patients without cancer incurred higher cumulative costs than CPD and CVD patients at all yearly time points, casting doubt on this explanation. A notable aspect of the results was the high magnitude of eight-year excess costs in the

no-comorbidity group, both relative to non-cancer controls and in absolute monetary terms. This remained, and in fact increased, after adjustment for confounders. Contributing factors may have been the lower base level of healthcare use in the no-comorbidity group combined with higher survival. Other studies have found that younger patients accumulate higher costs due to receiving more aggressive treatment [105]. However, phase-of-care costs for cancer patients in the no-comorbidity group point against this explanation, as costs were only notably higher in the end-of-life phase, while I would have expected them also to be higher in the initial phase and possibly the continuing one. On the other hand, if differences in treatment led to increased survival, the phase-of-care trajectories would support this explanation, as the charts indicate patients without comorbidities spent more time in the end-of-life phase.

Potential Bias

Adjustment for confounders did not significantly alter the direction of costs, and magnitudes were not notably affected. Residual confounding cannot be ruled out due to the lack of variables on additional confounders such as ethnicity and unreported comorbidities. Pre-diagnosis records could have provided proxy information on unreported comorbidities, however the phase-of-care trajectories in Figure 5.5 suggest little difference in pre-diagnosis costs between exposure cohorts, except in the no-comorbidity group. Additionally, differences in pre-diagnosis costs between exposure groups could have been related to cancer, making pre-diagnosis costs unsuitable as a proxy for pre-existing comorbidities. A likely contributing factor to the negative directions of all eight-year excess LTC-specific costs was the lower survival of cancer patients. Additional contributors may have been misattribution of costs to cancer and cancellation or postponement of LTC-related treatments for more urgent cancer-related ones. High counts of zero costs can complicate the estimation of healthcare costs [133], which prompted me to assess the robustness of estimations using two-part methods. The equality of means in GLM and two-part model estimations for LTC-specific costs supports these being unbiased, although confidence intervals were wider in two-part estimations because of larger standard errors. Despite the larger standard errors, the results for all cost coefficients and margins were significant at the 5% level.

Comparison to Other Studies

Comparison with other studies is challenging due to differences in study goals, populations and methods. Inpatient costs for cancer patients have been found to rise

with increased comorbidity [148] while another study found comorbidities in COPD patients associated with higher costs [202]. Other studies found limited associations between cancer and elevated costs for patients with chronic conditions [203] and [151], with the latter study finding both positive and negative associations with excess costs dependent on the comorbidity. A similar lack of association was replicated in a study that examined the effect of adjustment for comorbidities on the inpatient costs of multiple diseases [208], although the diseases examined did not include cancer. Differences in methodologies, populations and health systems may have contributed to the heterogeneity of results across studies. Yet even within studies the association of costs with the number of conditions can be condition-dependent. A possible explanation is that greater cost acquisition for multiple conditions is offset by diminished survival. The results of Blakely (2019) [151] support this explanation. While some combinations of conditions in this study were superadditive, with costs over and above the sums of the conditions, those with cancer showed limited interaction effects, and when cancer was combined with CVD, diabetes and LLK (lung, liver, kidney) the interaction effects were negative. A problem with generalising results from this analysis is that little information was given about how censoring was treated for patients who died, and the method of estimating costs for comorbidities was not fully described. The inclusion of primary care costs further decreases comparability with my findings. Despite these drawbacks, the results provide evidence that cancer interactions with LTCs may produce negative associations with costs despite other findings showing positive associations between costs and the number of comorbidities. Other explanations for the heterogeneity of associations may lie in the complexity of detection and subsequent treatment pathways. While chronic illness may increase the likelihood of cancer [6] it also increases exposure to the health system which may lead to earlier detection [118]. The stage of detection, in addition to directly affecting costs, may also interact with the comorbidity. The severity and type of comorbidity may also influence the treatment given, for example, severe COPD can make lung cancer inoperable [118], further affecting costs. The complexity of such interactions makes the interpretation of results challenging. Although later stage has been found to increase costs for cancer patients [135], the relationship may not be linear, and shorter survival of stage IV cancers may lower costs compared to stage III [179].

5.4.3 Strengths

The analysis benefited from a large sample with detailed individual-level data derived from patient healthcare records, enabling me to measure costs eight years beyond the

cancer diagnosis period and several years before it. Comparison with a cohort without comorbidities allowed measurement of the association between cancer and costs for multiple LTC groups. Yearly and phase-of-care trajectories highlighted different aspects of the cohort trajectories, providing new insights into the dynamics of costs over time.

5.4.4 Limitations

The analysis also had limitations. While residual matching caused by sampling from a larger matched cohort provided approximate similarity across exposure groups, it undermined the assumption of independence of observations, prompting methods that accounted for clustering of errors. An ideal design would have incorporated matched cancer and non-cancer groups within each LTC group, but this was not possible given the available dataset. Instead, I used the available data and adjusted for confounders using multivariable regression. While the use of whole-of-population administrative data provided a large sample size, using data not created to answer the research questions has downsides, such as unmeasured confounding caused by the omission of variables relevant to the study and bias due to changes in the coding of variables and subject eligibility over time [120]. Under-recording of comorbidities is a known issue in administrative data [118] that has been found to some extent in the SMR01 dataset [161, 205]. This could have led to both selection bias and confounding as SMR01 records provided data for inclusion in the LTC groups and for comorbidities in regression models. Attribution for LTC-specific costs came from the main condition variable rather than additional conditions. Under-reporting has also been found for the main condition in SMR01, meaning that total LTC-specific costs could be approximately 15% higher than reported. Misclassification of conditions has also been observed in SMR01 [161] but is more of an issue for less common conditions than those I studied, with diabetes and CVD having reported accuracy of 97% and COPD 99%. Additionally, diabetes is unlikely to be reported as the main condition for associated morbidities such as bone and urinary tract infections [209], meaning that diabetes-related costs were likely to be higher than reported.

Variables

The lack of reporting of the main condition in the outpatient records of SMR00, combined with the difficulties in attributing prescription records to conditions, led me to describing only inpatient costs. While inpatient costs are believed to account for the majority of healthcare costs [172], over long periods other services such as social care,

prescriptions, primary care and outpatients may incur substantial costs. This study, therefore, should not be taken as a full account of healthcare costs.

5.4.5 Generalisability

Issues concerning the external validity of costing studies were discussed in Section 3.4 and Section 4.4.4 and are also relevant here. Health factors specific to the Scottish population may have more bearing on this study due to the high prevalence of chronic conditions in Scotland compared to other western European countries [51, 210]. As previously discussed in Section 2.3.10, studies of healthcare costs typically have low external validity due to divergent methodologies, research questions, timescales, populations and health systems. The use of cost ratios rather than absolute monetary units should improve comparability for excess or attributable costs, but a cost ratio may be less straightforward for policymakers to interpret than a single monetary figure, while also being highly sensitive to the magnitude of the base level. This study used both types of measure but the monetary units should be interpreted cautiously, and overall patterns of results are more likely to be generalisable than particular outcome measures [151]. Differences across cancer and LTC groups may have greater external validity, as these ought to be less sensitive to differing methodologies, but the degree of bias inherent in a methodology may vary by condition. Similar considerations may apply to the trajectories of healthcare use over time. Here discount rates, perspective and censoring may cause variation across studies and across conditions.

5.4.6 Policy Implications and Future Research

Anticipating Costs

This analysis found that cancer lowered the inpatient costs specific to LTCs, while raising total inpatient costs only slightly. If cancer survival improves, the likely effect will be an increase in costs for patients with LTCs. Rising prevalence of LTCs is likely to increase overall healthcare costs further. Hence policymakers need to be aware that improvements in cancer treatment might increase rather than decrease the costs of chronic disease.

Screening

The group with the highest excess costs of cancer in this analysis had no measured morbidity prior to the cancer diagnosis, suggesting that their exposure to modifiable risk factors was low. The analysis in Chapter 3 suggested that early detection may

reduce costs. Further study could examine the impact of screening for younger people in good health, to investigate the potential for cost reduction based on early detection for this group. Widespread screening of younger people in good health is likely to entail considerable burdens with current technology, but these may be minimised by new technologies like those being developed for multi-cancer early detection (MCED) [211]. The high excess costs for people without morbidity prior to cancer suggest a potential for substantial cost savings, however this could depend on the cost of the treatments and the number of false positives requiring additional investigation.

Other Considerations

Attributing costs to specific conditions is not straightforward [96]. However, better reporting of conditions in SMR records could improve the accuracy of costing studies. On the other hand, the similarity of results for CPD and CVD highlights the interconnectedness of chronic diseases and validates approaches that focus on underlying risk factors rather than particular diseases. Although inpatient care is a major contributor, a complete picture of healthcare costs would include other services. Physical multimorbidity is associated with poorer mental health, and mental health problems can incur substantial costs [188]. Improved survival might also increase the burden on social care services and working-age benefits. Future research could examine the wider impact on health services of combinations of conditions and their underlying risk factors.

5.4.7 Conclusion

This investigation added depth to the previous analyses by measuring inpatient costs for patients with and without long-term conditions (LTCs), providing novel insights into the long-term costs of cancer. As cancer and LTCs increase in prevalence, this information will be of increasing interest to healthcare professionals, policymakers and health economists. The overall thesis aim was to examine the wider societal costs of cancer, which led to Objective 6, the aim of which was to better understand the relationship with employment. Information on employment was not recorded in SMR data, however, a goal of the thesis was to explore the linkage of administrative data to answer the research questions. To this end, efforts were made to link data from the Scottish Census but these efforts were unsuccessful. Hence an alternative datasource had to be found. The data used and the resulting analysis are described in Chapter 6.

6 Long-Term Employment Outcomes for Cancer Survivors in UKHLS

6.1 Introduction

The previous analyses measured healthcare costs, derived from linked administrative data, in the Scottish population. As the overall thesis aim was to explore the wider societal costs, I carried out an additional analysis on the association of cancer with employment. Although my interest was in the Scottish population, I was unable to obtain suitable data for this population specifically. Hence the analysis detailed in this chapter used data sampled from the whole UK.

As described in Section 2.3.3, a notable component of disease costs is productivity losses, which represent lost output arising from mortality and diminished work ability, both paid and unpaid. Losses in paid output can take several forms, such as absenteeism, when employees take time off work, presenteeism, which is lowered productivity during work hours, and leaving the workplace entirely, for example into unemployment or early retirement. The inclusion of productivity costs in costing studies is not universal [84], however, excluding effects on productivity is likely to underestimate the overall burden of a disease. Measuring productivity losses presents multiple challenges; a key one being the acquisition of informative data. Routine healthcare data typically do not include variables on productivity, however, linkage to public datasets such as unemployment and tax records could provide valuable insights into changes in income, employment and other productivity-related factors. A goal of this thesis was to link public datasets to gain a fuller understanding of the costs of disease. To that end, the feasibility of linking earnings data from HM Revenue and Customs (HMRC) was explored, however it became apparent that accessing this data would be challenging, and that even the best-case timeline for access would extend beyond the project's end. The alternative chosen was to link data from the Scottish Census with routine health data from the NHS, using probabilistic methods rather than the unique CHI number used for NHS records. Although the linkage was eventually carried out, delays in access compounded with the Covid crisis made analysing the data impossible. In the meantime, an alternative dataset was sought, and the publicly accessible UK Household Longitudinal Study (UKHLS) dataset was found to be suitable. While this dataset used a smaller sample than the Census dataset, was taken from the UK population rather than Scotland specifically, and lacked the richness of health variables contained in NHS datasets, it did have the advantage of containing many socioeconomic variables recorded longitudinally, which enabled analysis of how outcomes related to employment and productivity changed over time after a cancer diagnosis. The results provide novel insights into the socioeconomic burdens of cancer survivors in the UK.

6.2 Background

The UK has over two million cancer survivors, and with growth rates of 3% this number is expected to increase by one million per decade until 2040 [212]. Approximately half of cancer patients survive for at least 10 years and survival is highest for those aged 15–40 [4]. Although cancer is associated with ageing, half of cancers affect the working-age population [213]. The official UK retirement age for pension eligibility is set to increase [214], which is likely to create more working-age cancer survivors. Much research on the effects of cancer on employment has been carried out in the US, where health insurance claims provide accessible data on employment after illness, and where productivity costs are more commonly included in costing studies. Less is known about outcomes in the UK, and the differences in health and social security systems make inferences from US results problematic. Results from other European countries may provide more relevant references. Cross-country evidence indicates that cancer is associated with a higher risk of unemployment and early retirement; a meta-analysis of 36 studies found that cancer survivors were 37% more likely to be unemployed than healthy controls [213], although the risk varied between different types of cancer. One-third of the respondents (33%) in a UK survey stopped working, while other respondents reduced their hours or took unpaid leave [93]. Other studies also found reduced work hours [215, 216]. The largest impact on employment has been observed in the six months after diagnosis, when most treatment usually occurs [217, 218, 219, 220]. Beyond this period, many survivors return to work but may subsequently drop out of employment [221]. Several studies show a persistent negative association between a cancer diagnosis and working beyond six months [222, 223, 224, 225, 226, 219, 227], with a recent study finding an association at ten years after diagnosis [73]. However, other studies found little or no long-term effect on employment outcomes [228, 229]. Previous study has tended to focus on breast cancer in women, while outcomes for males, immigrants, and individuals with low socioeconomic status are less well known [230]. These studies and other analyses results discussed in this Chapter were identified using the literature search in Appendix C, with the focus being on identifying systematic reviews related to the likelihood of non-employment and unemployment. Additional studies were identified using Google Scholar, Web of Science, Cochrane Reviews and expert knowledge. The results are summarised in Table 6.1.

Table 6.1.: Summary of literature results of the association between cancer and employment and other outcomes related to in-work productivity

Study	Year	Currency	Population	Methods	Main results
Laudicella et al.	2016	UK sterling	NHS England patients	9 year incidence + 5 year prevalence	Incidence costs in the first year of diagnosis noticeably higher in patients age 18-64 than age >= 65 across all examined cancers. A lower stage diagnosis is associated with larger cost savings for colorectal and breast cancer.
Banegas et al.	2018	US dollars (USD)	45,522 cancer and 314,887 controls using us health plan data	Total and net costs using phase of care approach. 1 year and 5 year costs reported	Net costs were consistently highest for lung cancer and lowest for prostate cancer. Net costs were higher across all cancer sites for patients aged <65 years than those aged ≥65 years. Medical care costs for all cancers increased with advanced stage at diagnosis.
Yabroff et al.	2008	USD	718,907 cancer and 1,623,651 non-cancer us medicare patients (65+)	Net costs by phase of care using survival data to give 5 year costs	Mean net costs of care were highest in the initial and last year of life phases and lowest in the continuing phase. Mean 5-year net costs varied widely, from less than $20 000 for patients with breast cancer or melanoma of the skin to more than $40 000 for patients with brain or other nervous system, esophageal, gastric, or ovarian cancers or lymphoma.
Kutikova et al.	2005	USD	2040 lung cancer patients who were employees of large corporations in the US	Retrospective case control, inpatient, outpatient and drug costs, 2 years from diagnosis	Regression-adjusted mean monthly total costs were US dollar 6520 for patients versus US dollar 339 for controls (P < 0.0001), and overall costs across the study period (from diagnosis to death or maximum of 2 years) were US dollar 45,897 for patients and US dollar 2907 for controls (P < 0.0001). The main cost drivers were hospitalization (49.0% of costs) and outpatient office visits (35.2% of costs).
Pisu et al.	2018	USD	Multiple	Review of studies	For all payers combined, costs for cancers like breast, prostate, colorectal, and lung cancers were $20,000 to $100,000 in the initial phase, $1000 to $30,000 annually in the continuing phase, and >=$60,000 in the end-of-life phase.
Blakely et al.	2015	NZ dollars	New zealand cancer registry	Gamma regression on expected costs	From $5,000 (melanoma) to $66,000 (bone and connective tissue).
Lang et al.	2014	USD	Seer medicare stage IV breast cancer patients	Matched 1:1 retrospective analysis	Higher resource use in all areas except oral prescriptions.
Hanchate et al.	2010	USD	452 cancer patients and, 1,656 matched controls from various locations in the US	Prospective five-year longitudinal comparison of cases and matched controls	Breast cancer survivors' health care use and disease burden return to pre-diagnosis levels after one year but greater use of outpatient care persists at least five years.
Khanna et al.	2011	USD	West Virginia Medicaid administrative claims data for women recipients 21-64 years of age	Matched controls analysis	All-cause healthcare costs significantly higher for breast cancer patients than controls ($16,345 vs. $13,027, p<0.001).
Song et al.	2011	USD	6,675 patients with colorectal cancer matched to patients without cancer	Retrospective study using national claims database	Total monthly costs were $14,585, driven by higher inpatient care ($7,546) and outpatient care ($6,749).
Chang et al.	2004	USD	New diagnoses of one of seven types of cancer (n = 12,709). Controls without cancer were matched at a 3:1 ratio by demographics	Retrospective matched-cohort control analysis	Mean monthly costs ranged from 2,187 dollars for prostate cancer to 7,616 dollars for pancreatic cancer, most often driven by hospitalization. Costs for controls were 329 dollars per month.
Jayadevappa et al.	2005	USD	120 prostate cancer patients and 240 men without cancer, matched by age and race	Retrospective cohort control study using regression models	The incremental cost of prostate cancer was 1.30 times higher than controls.

USD = United States dollars, NZ = New Zealand.

6.2.1 Factors Influencing Return to Work

The likelihood of returning to work has multiple influences across three main themes: person factors, employment factors and wider contextual factors [230]. Factors known to influence return to work include type and stage of cancer [231, 232], type of employment [231, 232, 233], satisfaction with pre-cancer employment [234], and amount of support in the work environment [231, 220]. A study of breast cancer survivors found that most who became out of work due to cancer made the decision to leave work rather than suffer discrimination due to the cancer [224]. Additional factors known to affect return to work are social support from occupational health services, social factors at work, and having chronic diseases [232]. Cancer survivors may continue to work despite reporting physical and mental health issues [217], yet health reasons are likely to be a greater factor in non-employment for cancer survivors than those without a cancer history [213]. One quarter (25%) of employed cancer survivors reported that cancer interfered with physical tasks and 14% reported that it interfered with mental tasks [235]. Cancer-related burdens can last far beyond treatment [236, 237], yet their effect on long-term employment has received little study, partly because

inconsistency in terminology can make comparisons problematic. Burdens commonly reported are listed below.

Fatigue and Sleep Disorders

Cancer-related fatigue (CRF) is reported to be the most common side effect of cancer treatments, with up to 40% of patients suffering what has been described as *a subjective sense of exhaustion related to cancer and treatment that is non-proportional to recent activity* [238]. Persistent fatigue has been reported present in one quarter to one-third of cancer survivors at more than ten years after diagnosis [239]. Fatigue levels have been found to influence return to work [240]. However, due to the older age of cancer survivors, the degree to which fatigue can be attributed to the cancer diagnosis is difficult to determine [241]. A study found that survivors of stomach cancer experienced more fatigue in performing both household and employed work, and reported reduced working hours and reduced likelihood of working [242]. Problems with sleep are another common problem, with insomnia being the main complaint and hypersomnolence and obstructive sleep apnea also being observed [243]. Sleep disturbances may be related to fatigue and also to absenteeism and reduced work capacity [244].

Distress and Pain

Distress in cancer patients is a multifactorial experience that may be psychological, physical, social or spiritual, and can be distinguished from psychiatric disorders and generalised anxiety disorders [245]. Levels of anxiety are higher among cancer survivors than among the general population [246] while psychological distress is reported in up to half of cancer survivors long after treatment [247]. Physical pain is also common in survivors and can last up to 10 years from diagnosis, but the relationship with employment is unclear [248].

Cognitive Problems

Problems with mentation, concentration and memory are common cognitive complaints for cancer survivors that can arise as a result of chemotherapy [249]. It is estimated that 25-75% of chemotherapy recipients suffer from cancer-related cognitive dysfunction (CRCD), which can persist decades after treatment [250]. Cancer survivors can also experience accelerated ageing, which may be difficult to distinguish from normal ageing [251]. Sufferers may report difficulty coping with paid work and responsibility and may be moved to lower-paid positions or made redundant [250] even

though ability to work has been shown to improve over time [252].

Other Patient-Level Factors

Lower health-related quality of life (QoL) is associated with lower efficiency and higher absenteeism, though the observed associations were moderate [253]. Cancer sufferers may also suffer from other health problems. The average number of comorbid conditions for cancer survivors is five, both before and after diagnosis, with most conditions manifesting prior to the diagnosis [197]. Cancer survivorship can even enhance relationships and appreciation of life [249] which may deter survivors from returning to work [236]. Cosmetic effects of treatment can be a major issue for survivors [249], while levels of depression have been found to be higher in cancer patients than in the general population [246]. Furthermore, even if individuals regain full health, they may suffer employment "scarring" where previous spells of unemployment make present employment less likely and lower expected earnings [254].

Systemic and Social Factors

Differences in social security systems make cross-country comparisons challenging and the social security system itself can influence the likelihood of working for cancer survivors [236]. The duration of the association may be longer in countries with more generous social security systems [219, 213]. A meta-analysis found that survivors in the US had a 1.5 times higher risk of unemployment than those in Europe, however this risk disappeared when adjustment was made for diagnosis, age and background unemployment level [213]. The UK has unique institutions, in particular the NHS, and results from studies in other countries may not be generalisable to the UK context, particularly those from the US where the costs of treatment may be high and social support limited [255].

6.2.2 Aims and Objectives

Guidelines from NICE state that productivity costs should be excluded in technology evaluations but may be presented separately if their exclusion is an important aspect of the technology under assessment [256]. The exclusion of productivity costs from technology evaluations in the UK may be a factor in the relative scarcity of studies undertaken on employment outcomes for working-age cancer survivors. Methodological and practical issues make their incorporation into evaluations problematic, but understanding these outcomes can aid policymakers, support services, employers and health professionals in providing adequate support for survivors, and enhance

understanding of the overall costs of cancer to society. In particular, the long-term effects on employment of cancer survival are not well understood in the UK, but are likely to become more important as cancer survival improves. The aim of this study was to better understand the relationships between a cancer diagnosis and work-related outcomes for cancer survivors in the UK, with an emphasis on whether changes in outcomes persist beyond the treatment period. The specific objectives were as follows.

1. Measure the long-term association between a cancer diagnosis and the likelihood of being in paid work.
2. Measure the long-term associations between a cancer diagnosis and other work-related outcomes such as earnings, income and self-reported health.

6.3 Overall Methods and Data

The investigation was carried out as two analyses, each using data from the same dataset. Due to the distinctness of methods and samples, the analyses are presented separately, followed by a discussion of the results for both analyses. Analysis A used regression modelling on a cross-section of data while Analysis B used difference in differences (DiD) methods on data spanning multiple time periods. For the first analysis, a cross-sectional approach was attractive for a number of reasons. The health variables of interest had better coverage in wave 1, with fewer missing data and a simpler structure than in other waves. Of particular interest was the time-since-diagnosis variable, as this could be used to study outcomes at a greater temporal distance from the exposure than the survey follow-up allowed, but this variable had low completion (<50%) in subsequent waves. Furthermore, the coding of the health variables changed throughout the study, making a consistent measure of health unobtainable. Furthermore, restricting analysis to the first wave of data avoided the problem of attrition due to individuals dropping out over subsequent waves. Analysis B utilised the longitudinal aspect of the dataset by measuring outcomes as individuals moved through the survey period, using new diagnoses as exposures rather than the time-since-diagnosis variable in Analysis A. While longitudinal data suffer from attrition, they can increase statistical power and provide adjustment for unobservable confounders [257], while following individuals over time could provide evidence for a causal effect [258].

Data for both analyses were taken from the UK Household Longitudinal Study (UKHLS), a panel study primarily funded by the UK Economic and Social Research Council (ESRC), University of Essex and distributed by the UK Data Service, to aid

understanding of the long-term effects of social and economic change, including policy interventions. UKHLS was designed to be representative of the UK population and comprised approximately 100,000 individuals in total, taken during the period 2008–18 at the time of this analysis. I analysed the general population sample, which used stratified, clustered, equal probability sampling [259] to draw 47,520 addresses from Great Britain and 2,395 addresses from Northern Ireland for a total of 49,915 addresses in the UK [260]. Subjects were surveyed primarily through computer-aided personal interview (CAPI) [260] by trained interviewers, with other data collection via self-completion instruments, telephone interviews and online questionnaires.

The primary outcome under investigation was likelihood of employment vs non-employment. Other outcomes of interest were total income, earnings from employment, and hours worked. I also investigated physical and mental self-reported health as outcomes as these are believed to be associated with employment [261, 228, 262] and hence could provide support for the results. They were also of interest as indicators of the factors influencing return to work described in Section 6.2.1. More information on the self-reported health variables is given in Section 6.4.3.

6.4 Analysis A: Methods

6.4.1 Analysis A: Study Overview

Data from the cross-sectional survey were analysed using univariable and multivariable logistic and linear regression models. The exposed group was stratified by time to measure associations between cancer and employment outcomes at temporal distances from the diagnosis. UKHLS took addresses from all over Great Britain and Northern Ireland to build a representative sample of the UK population. Interviews were conducted via self-completion instruments, telephone interviews and online questionnaires. Data collection for wave 1 of UKHLS covered the period 2008–9 with a single interview for each individual in the study. Data for each particular wave in UKHLS were collected over a period of 24 months, hence waves do not correspond precisely to calendar years, though interviews for an individual were scheduled one year apart.

6.4.2 Analysis A: Participants

My interest was in working-age cancer survivors and working-age controls without a history of cancer. UKHLS comprises multiple waves of interviews over time, which

could increase the number of records. However, as the same individuals were followed over time, including multiple waves would not substantially increase the number of participants. Although additional participants entered the study in waves 2 and 3, others in wave 1 also dropped out. Furthermore, the time-since-diagnosis variable that I used to stratify the exposed group had high completion in wave 1 but low completion in subsequent waves. The numbers in wave 1 were considered sufficient to measure the associations of interest, and restricting analysis to a single wave avoided secular effects and attrition. Hence I restricted analysis to the first wave of the dataset.

All individuals were required to be 30–64 years of age for men and 30–59 years of age for women. I chose a lower bound of 30 years to allow sufficiently long time periods to have passed since the individual received the cancer diagnosis, as younger individuals may have received their cancer diagnosis at a very young age, and childhood cancers were not the main focus of study. Additionally, individuals below 30 years old can be assumed more likely to be in education, less likely to be in employment, and less likely to have a history of cancer, hence excluding these individuals reduced the likelihood of confounding due to age. The percentage of cancer patients under 30 years of age was relatively low at approximately 6% ($N = 38$). The upper age limits were chosen to ensure that the participants were of working age. Differences between upper age limits for men and women reflected government policy on pension eligibility for the study period.

Approximately 5% of the responses were obtained by proxy, meaning that information about these individuals came from a household member rather than the individual directly. Individuals with data obtained from proxy interviews were excluded because these records lacked information on health conditions. Individuals without information on the time since diagnosis were also excluded. This led to the following participant numbers:

Exposed (N=657): a cancer diagnosis at any time prior to the interview.

Controls (N=25,968): no history of cancer at the time of the interview.

6.4.3 Analysis A: Variables

Outcome Variables

Working: The primary outcome was the employment status of an individual at the time of the interview, restricted to two possible outcomes: working and not working.

To measure this, I created a single binary variable from the variables *jbhas* and *jboff* in the UKHLS dataset, in line with UKHLS guidelines. These variables related, respectively, to the questions *Can I just check, did you do any paid work last week - that is in the seven days ending last Sunday - either as an employee or self-employed?* and *Even though you weren't working did you have a job that you were away from last week?* Hence working individuals were defined as being in paid employment (full or part-time) or self-employment at the time of their interview, and included those on maternity leave or sick leave. Non-working individuals were all those outside paid employment, including the unemployed, retired, students, inactive and homemakers.

Additional outcomes were analysed to better understand the association between cancer and productivity. These included income, job hours and earnings for those in work, and self-reported health measures that incorporated health-related factors described in Section 6.2.1, that could affect the ability to work.

Hours worked per week: This continuous variable could take a value of zero for those not in employment, or positive values for those employed with possible fractional values. Individuals in UKHLS were asked the question *Thinking about your (main) job, how many hours, excluding overtime and meal breaks, are you expected to work in a normal week?*.

Income: To analyse income I used the continuous variable *fimnnet* in UKHLS, which measured monthly income net of taxes on earnings and national insurance contributions, and was constructed in UKHLS from the sum of six income components: labour income, miscellaneous income, private benefit income, investment income, pension income and social benefit income. In addition to positive and zero values, this variable could take negative values, believed to be due to loss-making investment income

Earnings: Earnings data were taken from the continuous variable *fimnlbgrs* in UKHLS, which measured gross monthly earnings from paid or self-employment. Earnings differed from income by only representing paid employment or self-employment and did not include tax deductions. This variable could take negative, positive and zero values.

SF-12 Physical and Mental Summary Scores: The Short Form 12 (SF-12), a condensed version of SF-36, is a standardised instrument of 12 questions that measure general health and well-being from a patient's perspective [263]. Two summary scores in the range [0,100] representing self-reported physical and mental health are derived

from the twelve answers. While not direct employment outcomes, the physical and mental summary scores incorporate information on factors related to work ability described in Section 6.2.1. I included these variables as outcomes to better understand the relationship between changes in health and changes in employment over time, and as support for evidence of associations with the employment outcomes. The summary scores of SF-12 can be considered measures of health-related quality of life (HRQoL) and are designed to have mean values of 50 across respondents and approximate normal distributions for large samples.

Exposure Variables

Cancer diagnosis at any time: In wave 1, interviewees were asked *Has a doctor or other health professional ever told you that you have any of these conditions?* Seventeen conditions were listed in wave 1 including cancer. A binary variable representing whether an individual had received a cancer diagnosis at any time in the past took values: 0 = no cancer diagnosis, or 1 = diagnosed with any type of cancer or malignancy at any time prior to the interview.

Cancer diagnosis stratified by time: Individuals were also asked *What age were you when you were first told you had <condition>?* I used the information from these responses to derive a categorical variable representing whether an individual had ever received a cancer diagnosis at one of the time periods described below. Using a variable with multiple levels was convenient for logistic regression, where the *no cancer* diagnosis is used as a base case from which to compare the odds ratios of other groups. While the categories would ideally reflect treatment and post-treatment phases after a diagnosis, the lack of reporting precision combined with possible inaccuracies of recall made this unattainable, and categories were chosen to balance statistical power with informative look-back while approximating the initial period of shock after diagnosis. Participants with childhood cancers were included in the >10 years group, a total of 8 individuals who had cancer before the age of 16. The categories were:

0 = no cancer diagnosis
1 = cancer diagnosis at 1–2 years prior to interview
2 = cancer diagnosis at 3–5 years prior to interview
3 = cancer diagnosis at 6-10 years prior to interview
4 = cancer diagnosis at >10 years prior to interview

Confounders and Other Variables

Female: A binary variable was used to represent the sex of respondents. This variable took values: 0 = male and 1 = female.

Age: Three age categories were created: 30–44 years, 45–54 years, 55–64 years, which were chosen to balance the numbers in categories while providing simple and informative groupings. For simplicity, the same coding was used for men and women, meaning that the upper category for women was 55–59 years. In addition, continuous age variables and the square of age were included in multivariable analyses to adjust for residual confounding. While the categorical variables were created for reporting sample characteristics, they were also included in regression models where model selection based on Bayesian information criterion (BIC) and Akaike information criterion (AIC) suggested their inclusion produced superior models.

Higher qualification: The highest education variable in the raw UKHLS dataset gave multiple categories of highest education attained. To account for low numbers in some categories for the exposed groups, the categories were aggregated to a binary variable with values 0 = no higher level qualification, 1 = higher level qualification (degree and above).

No parents working at age 14: As a proxy for childhood deprivation, I created a binary variable to record whether an interviewee's parents worked during the interviewee's childhood. In UKHLS, interviewees were asked *Thinking back to when you were 14 years old, was your father working at that time? And was your mother working when you were 14?* to give two variables with four possible values each, which were aggregated to a single binary variable with values: 0 = at least one parent working when individual was aged 14, 1 = no parents working when individual was aged 14. The raw variable in wave 1 contained substantial numbers of missing values. These were assigned values from the same individual in later waves of the longitudinal data, under the assumption that the values were fixed throughout the study.

Marital status: Four categories of the raw variable representing marital status were aggregated into a binary variable with values: 0 = not married, 1 = married.

Ethnicity: Eleven ethnic categories in the UKHLS dataset were aggregated into a binary variable with values: 0 = white UK ethnicity, 1 = other ethnicity.

Other health conditions: Seventeen health conditions (including cancer) were recorded in wave 1, and with each was included a supplementary time since diagnosis

variable. I created a binary variable for each health condition with values: 0 = no diagnosis, 1 = diagnosis at any time prior to interview. Additionally, I created variables for each health condition stratified by time using the same method as the cancer exposure with the same categories:

0 = no diagnosis
1 = diagnosis at 1–2 years before interview
2 = diagnosis at 3–5 years before interview
3 = diagnosis at 6–10 years before interview
4 = diagnosis at >10 years before interview

Job status: I used the variable *jbstat* in the sample characteristics table to provide details on the socioeconomic groupings of the sample. This variable was used for reporting of baseline characteristics but was excluded from the analyses due to low numbers in categories.

Sample weights: Cross-sectional and longitudinal sample weights were included in UKHLS, taking continuous values between zero and one. These were used in sensitivity analysis to explore the extent of bias in the sample.

Analysis A: Quantitative Variables

Age and the time since diagnosis were grouped as described above. It was expected *a priori* that employment likelihood and earnings would rise through adulthood, then decline during late middle-age, making a non-linear age term appropriate. AIC and BIC were used for model selection and on this basis, age and its quadratic were included in models but not higher powers. The number of hours worked per week was approximately normal for those working, but contained a large mass at zero representing those who were not working. Earnings also contained high zero mass for non-working individuals. I restricted analysis of earnings and work hours to those working, either self-employed or waged. Earnings and income showed considerable right-skew hence I took the natural logarithms to account for this [264]. However, both earnings and income contained negative values, and it was not certain that transformed values were more appropriate for regression modelling. Therefore I analysed both raw and transformed values. SF-12 physical and mental summary scores are designed to be normally distributed in large samples, with mean at 50. The distributions of the data under observation approximated these properties, though some left-skew was noted, hence these variables were used without adjustments.

6.4.4 Analysis A: Missing Data and Bias

Higher qualification had approximately 10% missing values. After recoding the categories to binary values, logistic regression was used as a single imputation method. The level of qualification was associated with employment and income-related outcomes; hence, it was important to adjust for this. Multinomial regression on the full list of categories was explored; however, the predictive accuracy of these models was lower while maintaining the full list of categories did not improve model selection based on BIC. Poor completion was also found in the *no parents working at age 14* variable. As this variable can be considered to be fixed throughout the follow-up, an approach similar to baseline observation carried forward (BOCF) single imputation was carried out, but in reverse—taking values from later longitudinal waves. The remaining missing values were predicted using logistic regression as a single imputation method. Other variables had full completion or <0.1% missing values.

Males and people with severely limiting long-term illnesses have been found to be under-represented in UKHLS data [265]. To reduce known biases the UKHLS dataset includes sample weights for each wave. Analyses using sample weights showed slightly higher association magnitudes with direction unchanged, and higher significance, suggesting the raw data had a downward bias on employment effects. I decided to report the main analyses without sample weights, to avoid overestimating associations and because weighting complicated the interpretation of descriptive data in relation to analysis results. Sensitivity analyses using sample weights were carried out and are described with results included as Appendices.

Health data and other data were self-reported by those interviewed and may therefore be prone to biases inherent in self-reported data. The use of objective health indicators, such as the diagnosis of a major disease, can minimise reporting bias, relative to subjective measures of self-reported health, but in UKHLS these health indicators too are self-reported and can be prone to recall bias. Inaccuracies in recall and also reverse causation can lead to biased associations between the exposure and outcome [266]. To avoid problems of reverse causation I decided to include only confounders that would have occurred before the cancer diagnosis, however, for some variables, such as highest qualification and marital status, this was not feasible with the data provided and I took the approach that reverse causation would cause less bias than excluding these confounders. With regard to comorbidities, it was technically possible to place the diagnosis of comorbidities relative to the cancer diagnosis; however, such temporal positioning was only relevant to the exposed group. Therefore,

I controlled for comorbidities in separate regression models by discarding the time categories and using only binary indicators. This approach also allowed me to compare the associations of different diseases with the outcomes to those of cancer.

Individuals who conducted interviews by proxy were excluded, as health variables were not available for these individuals. The use of a proxy interviewer was likely to be associated with poorer health, as it may have indicated that the interviewee was too sick to answer questions directly. This could have biased results in the likely direction of an underestimation of the association. Alternatively, it was possible that people in work had less time to participate in interviews, which could have led to overestimations of associations, as results for those out of work would take greater weight. Approximately 5% of the observations were made by proxy, totalling 1,812 individuals in wave 1.

6.4.5 Analysis A: Statistical Methods

Tables of characteristics for the cancer and control groups were produced, with and without stratification of the cancer cohort by time since diagnosis. Histograms of the outcomes were created and I also created charts of the outcomes with the cancer group stratified by time since diagnosis. Univariable regression analysis was performed to measure the association between a cancer diagnosis at any time and the described outcomes. The binary outcome *working* took a logistic model with logit function

$$logit(Pr(working_i = 1)) = ln\left(\frac{Pr(working_i = 1)}{1 - Pr(working_i = 1)}\right) \tag{6.1}$$

Giving the logistic regression model

$$Pr(working_i = 1) = \frac{e^{\alpha+\lambda cancer_i}}{1 + e^{\alpha+\lambda cancer_i}} \tag{6.2}$$

This takes a probabilistic interpretation to describe the likelihood that an individual i with cancer diagnosis will be working, with a constant α and coefficient λ to be estimated.

The continuous outcomes took univariate linear models with the particular outcome a linear function of cancer.

$$y_i = \alpha + \lambda cancer_i + u_i \tag{6.3}$$

Where y_i is a particular employment outcome for individual i, $alpha$ is a constant, λ is a coefficient to be estimated and u is a normally distributed random error term.

For the outcomes of earnings and hours worked, only working individuals were analysed. This prevented the outcome distributions having large numbers of zero values, which are statistically undesirable. Statistical models exist to deal with high zero counts, such as two-part models and GLMs, however these are less straightforward to estimate and interpret, while limiting observation to those in work added the additional aspect of analysing in-work outcomes. Adjustment for confounders was carried out using multivariable models. The multivariable logistic model for the binary outcome of working took the overall form

$$Pr(working_i = 1) = \frac{e^{\alpha + \lambda cancer_i + \gamma X_i}}{1 + e^{\alpha + \lambda cancer_i + \gamma X_i}} \tag{6.4}$$

Here X_i is a vector of observed and unobserved confounders for individual i and other symbols are as before. The multivariable model for continuous outcomes similarly extends the univariable linear model.

$$y_i = \alpha + \lambda cancer_i + \gamma X_i + u_i \tag{6.5}$$

Again, X_i is a vector of observed and unobserved confounders for individual i, γ is a vector of coefficients to be estimated and other symbols are as before.

Adjustment for comorbidities was complicated by the possibility of reverse causation, where the cancer diagnosis occurred before the comorbid condition and was a direct or indirect cause of it, potentially through the outcome as mediator. This could have caused biased estimates if the effect was attributed to the comorbidity rather than the cancer, or vice versa. While models incorporating interactions may have been possible, the number of interactions between the comorbidities and the stratified cancer exposures would have been impractically large. I therefore presented the main results for multivariable analysis without comorbidities. Results of models incorporating

comorbidities were presented as an additional analysis.

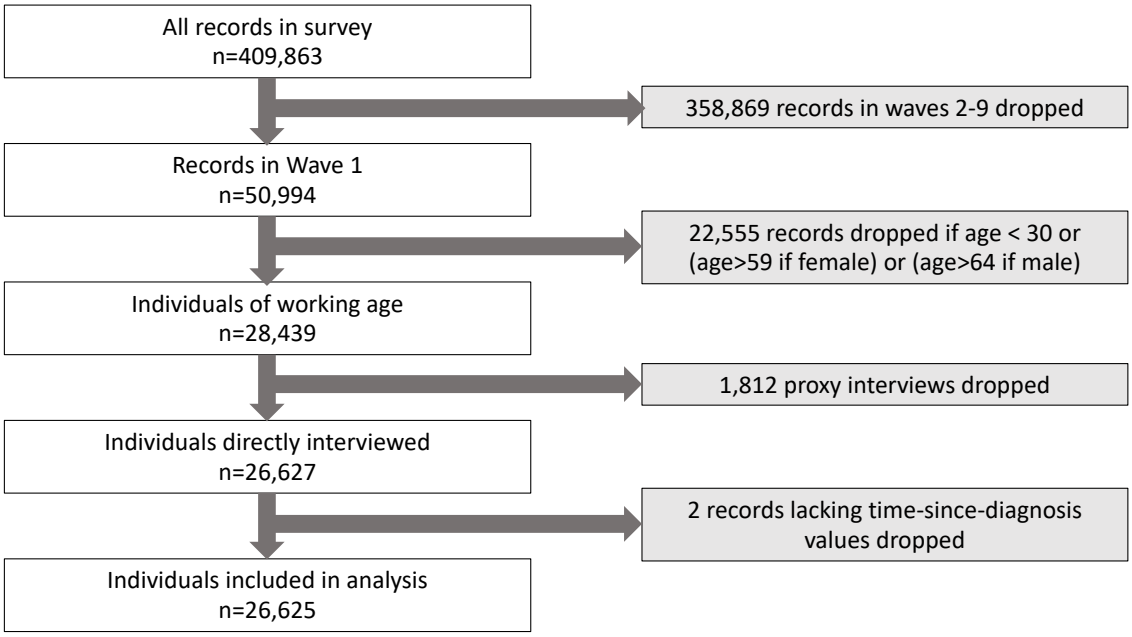
Coefficients for linear models with continuous outcomes were estimated using ordinary least squares (OLS). Those for logistic models with binary outcomes were estimated using maximum likelihood (ML) and reported as odds ratios. Model selection was made on the basis of the BIC and AIC, with lower values for these being preferred. Standard errors were reported in all analyses with robust methods used throughout. R-squared values were given for linear models as an indication of explanatory power, while the area under the receiver operating curve (AUC) was shown for logistic models. For linear models, the normality of residuals was assessed visually using Q-Q plots. The main results presented did not utilise sample weights in their estimation, however additional analyses were carried out to test for the possibility of sample bias, its direction and magnitude, using sample weights. In all statistical tests 5% significance levels were used and I reported estimates with 95% confidence intervals. All analyses were carried out in Stata 15.1.

6.5 Analysis A: Results

6.5.1 Analysis A: Participants

The derivation of participant numbers from the UKHLS dataset is given in Figure 6.1. A total of 409,863 records across nine waves were present in the UKHLS dataset at the time of the analysis. These were filtered to 26,625 unique records in total, comprising 657 people with cancer and 25,968 people without cancer.

Figure 6.1.: Flow chart of study participant numbers in Analysis A



6.5.2 Analysis A: Descriptive Data

Table 6.2 shows the characteristics of the cancer and non-cancer groups with additional detail on the exposure groups given in Table 6.3. A higher proportion of women was present in the cancer group, possibly reflecting the relatively high prevalence of breast cancer among working-age women. Cancer patients were older on average with a considerably lower proportion in the 30–44 age group. Marital status was not substantially different; however, considering the age discrepancies between the groups, this similarity may be misleading. The proportion of individuals in paid employment was lower in the cancer group, with higher proportions in retirement and long-term sick or disabled. However, the proportions of unemployment were lower in the cancer group than in the control group. The cancer group had a slightly lower proportion without higher level qualification and a slightly higher proportion with no parents working at age 14. Cancer proportions were higher in UK white individuals than in other groups. Self-reported physical health, measured in SF-12, was notably lower in the cancer group, however self-reported mental health showed a smaller difference. All health conditions other than congestive heart failure were represented more strongly in the cancer group than in the control group; however, this may have been related to the higher age of cancer patients.

Table 6.2.: Sample characteristics of the cancer and non-cancer participants in Analysis A

	Cancer N=657	No cancer N=25,968	Total N=26,625
Time from cancer diagnosis			
1-2 years	127 (19.3%)	NA	127 (0.5%)
3-5 years	164 (25.0%)	NA	164 (0.6%)
6-10 years	154 (23.4%)	NA	154 (0.6%)
>10 years	212 (32.3%)	NA	212 (0.8%)
Age in years (mean, sd)	50.3 (8.3)	44.8 (9.2)	44.9 (9.2)
Age group			
30-44 years old	173 (26.3%)	13,271 (51.1%)	13,444 (50.5%)
44-54 years old	262 (39.9%)	7,915 (30.5%)	8,177 (30.7%)
55-64 years old	222 (33.8%)	4,782 (18.4%)	5,004 (18.8%)
Sex			
Male	257 (39.1%)	12,184 (46.9%)	12,441 (46.7%)
Female	400 (60.9%)	13,784 (53.1%)	14,184 (53.3%)
Marital status			
Unmarried or separated	257 (39.2%)	10,050 (38.7%)	10,307 (38.7%)
Married or civil partners	399 (60.8%)	15,912 (61.3%)	16,311 (61.3%)
Employment status			
Not working	244 (37.1%)	7,023 (27.1%)	7,267 (27.3%)
Working	413 (62.9%)	18,938 (72.9%)	19,351 (72.7%)
Current economic activity			
Self employed	54 (8.2%)	2,724 (10.5%)	2,778 (10.4%)
Paid employment	348 (53.0%)	15,892 (61.2%)	16,240 (61.0%)
Unemployed	40 (6.1%)	1,890 (7.3%)	1,930 (7.2%)
Retired	48 (7.3%)	889 (3.4%)	937 (3.5%)
Long-term sick or disabled	113 (17.2%)	1,494 (5.8%)	1,607 (6.0%)
Doing something else	54 (8.2%)	3,077 (11.9%)	3,131 (11.8%)
Ethnicity			
UK white	585 (89.2%)	19,125 (73.7%)	19,710 (74.1%)
Non UK white	71 (10.8%)	6,828 (26.3%)	6,899 (25.9%)
Higher level qualification			
No higher qualification	399 (60.8%)	15,212 (58.7%)	15,611 (58.8%)
Higher level qualification	257 (39.2%)	10,689 (41.3%)	10,946 (41.2%)
*Hours normally worked per week (mean, sd)	32.8 (10.4)	33.6 (10.7)	33.6 (10.7)
*Monthly gross earnings (mean, sd)	1,245 (1,555)	1,482 (1,648)	1,476 (1,646)
Total gross monthly income (mean, sd)	1,384 (1,259)	1,475 (1,490)	1,472 (1,485)
SF-12 physical component summary (mean, sd)	43.4 (13.8)	50.8 (10.5)	50.6 (10.6)
SF-12 mental component summary (mean, sd)	48.0 (11.4)	49.8 (10.2)	49.8 (10.2)
No parents working at age 14			
At least one parent working	614 (93.6%)	23,913 (92.2%)	24,527 (92.2%)
No parents working	42 (6.4%)	2,020 (7.8%)	2,062 (7.8%)
Asthma	105 (16.0%)	3,161 (12.2%)	3,266 (12.3%)
Arthritis	122 (18.6%)	2,700 (10.4%)	2,822 (10.6%)
Congestive heart failure	1 (0.2%)	91 (0.4%)	92 (0.3%)
Coronary heart disease	16 (2.4%)	273 (1.1%)	289 (1.1%)
Angina	14 (2.1%)	367 (1.4%)	381 (1.4%)
Heart attack or myocardial infarction	19 (2.9%)	327 (1.3%)	346 (1.3%)
Stroke	16 (2.4%)	264 (1.0%)	280 (1.1%)
Emphysema	12 (1.8%)	105 (0.4%)	117 (0.4%)
Hyperthyroidism or an over-active thyroid	8 (1.2%)	225 (0.9%)	233 (0.9%)
Hypothyroidism or an under-active thyroid	36 (5.5%)	630 (2.4%)	666 (2.5%)
Chronic bronchitis	41 (6.2%)	464 (1.8%)	505 (1.9%)
Liver condition of any kind	20 (3.0%)	355 (1.4%)	375 (1.4%)
Diabetes	45 (6.8%)	1,320 (5.1%)	1,365 (5.1%)
Epilepsy	16 (2.4%)	314 (1.2%)	330 (1.2%)
High blood pressure	142 (21.6%)	3,958 (15.2%)	4,100 (15.4%)
Clinical depression	96 (14.6%)	2,116 (8.1%)	2,212 (8.3%)

* Calculated for individuals in work only. Earnings are for paid salaried and self employment.

Notes: SF-12 = Short-Form 12, sd = standard deviation. Indents indicate multiple categories of a single variables. Binary and continuous variables are unindented.

Table 6.3.: Sample characteristics of the cancer and non-cancer participants by exposure group in Analysis A

	Time since cancer or malignancy diagnosis					
	No cancer N=25,968	**1-2 years** N=127	**3-5 years** N=164	**6-10 years** N=154	**>10 years** N=212	**All records** N=26,625
Age in years (mean, SD)	44.8 (9.2)	51.1 (9.1)	50.8 (8.0)	50.2 (7.9)	49.4 (8.3)	44.9 (9.2)
Age group						
30-44 years	13,271 (51.1%)	34 (26.8%)	41 (25.0%)	34 (22.1%)	64 (30.2%)	13,444 (50.5%)
44-54 years	7,915 (30.5%)	42 (33.1%)	63 (38.4%)	73 (47.4%)	84 (39.6%)	8,177 (30.7%)
55-64 years	4,782 (18.4%)	51 (40.2%)	60 (36.6%)	47 (30.5%)	64 (30.2%)	5,004 (18.8%)
Sex						
Male	12,184 (46.9%)	60 (47.2%)	62 (37.8%)	58 (37.7%)	77 (36.3%)	12,441 (46.7%)
Female	13,784 (53.1%)	67 (52.8%)	102 (62.2%)	96 (62.3%)	135 (63.7%)	14,184 (53.3%)
Marital status						
Unmarried or separated	10,050 (38.7%)	54 (42.5%)	66 (40.5%)	52 (33.8%)	85 (40.1%)	10,307 (38.7%)
Married or civil partners	15,912 (61.3%)	73 (57.5%)	97 (59.5%)	102 (66.2%)	127 (59.9%)	16,311 (61.3%)
Employment status						
Not working	7,023 (27.1%)	51 (40.2%)	68 (41.5%)	53 (34.4%)	72 (34.0%)	7,267 (27.3%)
Working	18,938 (72.9%)	76 (59.8%)	96 (58.5%)	101 (65.6%)	140 (66.0%)	19,351 (72.7%)
Current economic activity						
Self employed	2,724 (10.5%)	5 (3.9%)	13 (7.9%)	12 (7.8%)	24 (11.3%)	2,778 (10.4%)
Paid employment(ft/pt)	15,892 (61.2%)	64 (50.4%)	82 (50.0%)	88 (57.1%)	114 (53.8%)	16,240 (61.0%)
Unemployed	1,890 (7.3%)	8 (6.3%)	11 (6.7%)	8 (5.2%)	13 (6.1%)	1,930 (7.2%)
Retired	889 (3.4%)	9 (7.1%)	11 (6.7%)	11 (7.1%)	17 (8.0%)	937 (3.5%)
LT sick or disabled	1,494 (5.8%)	30 (23.6%)	37 (22.6%)	22 (14.3%)	24 (11.3%)	1,607 (6.0%)
Doing something else	3,077 (11.9%)	11 (8.7%)	10 (6.1%)	13 (8.4%)	20 (9.4%)	3,131 (11.8%)
Ethnicity						
UK white	19,125 (73.7%)	110 (86.6%)	144 (88.3%)	139 (90.3%)	192 (90.6%)	19,710 (74.1%)
Non UK white	6,828 (26.3%)	17 (13.4%)	19 (11.7%)	15 (9.7%)	20 (9.4%)	6,899 (25.9%)
Higher level qualification						
No higher qualification	15,212 (58.7%)	79 (62.2%)	97 (59.1%)	95 (61.7%)	128 (60.7%)	15,611 (58.8%)
Higher level qualification	10,689 (41.3%)	48 (37.8%)	67 (40.9%)	59 (38.3%)	83 (39.3%)	10,946 (41.2%)
Hours normally worked per week (mean, SD)	33.6 (10.7)	33.8 (10.3)	31.3 (11.2)	32.2 (10.4)	33.9 (9.9)	33.6 (10.7)
Monthly labour income gross (mean, SD)	1,482 (1,648)	1,325 (1,631)	1,087 (1,528)	1,202 (1,428)	1,349 (1,617)	1,476 (1,646)
Total monthly income gross (mean, SD)	1,475 (1,490)	1,358 (1,017)	1,268 (1,003)	1,452 (1,418)	1,441 (1,434)	1,472 (1,485)
SF-12 physical component summary (mean, SD)	50.8 (10.5)	40.3 (14.3)	42.1 (14.2)	44.1 (13.3)	45.7 (13.0)	50.6 (10.6)
SF-12 mental component summary (mean, SD)	49.8 (10.2)	48.0 (11.5)	48.7 (11.6)	49.2 (10.3)	46.7 (12.0)	49.8 (10.2)
No parents working at age 14						
At least one parent working	23,913 (92.2%)	117 (92.1%)	153 (93.3%)	147 (95.5%)	197 (93.4%)	24,527 (92.2%)
No parents working	2,020 (7.8%)	10 (7.9%)	11 (6.7%)	7 (4.5%)	14 (6.6%)	2,062 (7.8%)
Asthma	3,161 (12.2%)	19 (15.0%)	14 (8.5%)	25 (16.2%)	47 (22.2%)	3,266 (12.3%)
Arthritis	2,700 (10.4%)	24 (18.9%)	34 (20.7%)	26 (16.9%)	38 (17.9%)	2,822 (10.6%)
Congestive heart failure	91 (0.4%)	0 (0.0%)	0 (0.0%)	1 (0.6%)	0 (0.0%)	92 (0.3%)
Coronary heart disease	273 (1.1%)	2 (1.6%)	8 (4.9%)	3 (1.9%)	3 (1.4%)	289 (1.1%)
Angina	367 (1.4%)	4 (3.1%)	5 (3.0%)	4 (2.6%)	1 (0.5%)	381 (1.4%)
Heart attack or myocardial infarction	327 (1.3%)	5 (3.9%)	6 (3.7%)	4 (2.6%)	4 (1.9%)	346 (1.3%)
Stroke	264 (1.0%)	1 (0.8%)	5 (3.0%)	2 (1.3%)	8 (3.8%)	280 (1.1%)
Emphysema	105 (0.4%)	3 (2.4%)	4 (2.4%)	1 (0.6%)	4 (1.9%)	117 (0.4%)
Hyperthyroidism or an over-active thyroid	225 (0.9%)	1 (0.8%)	3 (1.8%)	2 (1.3%)	2 (0.9%)	233 (0.9%)
Hypothyroidism or an under-active thyroid	630 (2.4%)	1 (0.8%)	12 (7.3%)	13 (8.4%)	10 (4.7%)	666 (2.5%)
Chronic bronchitis	464 (1.8%)	8 (6.3%)	8 (4.9%)	10 (6.5%)	15 (7.1%)	505 (1.9%)
Any kind of liver condition	355 (1.4%)	3 (2.4%)	3 (1.8%)	10 (6.5%)	4 (1.9%)	375 (1.4%)
Diabetes	1,320 (5.1%)	8 (6.3%)	12 (7.3%)	14 (9.1%)	11 (5.2%)	1,365 (5.1%)
Epilepsy	314 (1.2%)	1 (0.8%)	4 (2.4%)	4 (2.6%)	7 (3.3%)	330 (1.2%)
High blood pressure	3,958 (15.2%)	22 (17.3%)	36 (22.0%)	40 (26.0%)	44 (20.8%)	4,100 (15.4%)
Clinical depression	2,116 (8.1%)	13 (10.2%)	27 (16.5%)	20 (13.0%)	36 (17.0%)	2,212 (8.3%)

Notes: SF12 = Short-Form 12, sd = standard deviation. Indents indicate multiple categories of a single variables. Binary and continuous variables are unindented.

6.5.3 Analysis A: Outcome Data

Histograms of outcomes for cancer and non-cancer groups are shown in figures 6.2–6.7. Figure 6.2 indicates that the cancer group had a lower proportion of working, while the histogram shapes were not substantially different between exposure groups for hours worked, earnings and income as seen in figures 6.3–6.5. SF-12 summary scores shown in figures 6.6–6.7 had peaks at around 50 as expected, with the peak for

SF-12 physical having notably lower density for the cancer group, while left-skew was also present for both the cancer and no-cancer groups.

Figure 6.2.: Histograms of working vs not working for participants with and without cancer in Analysis A

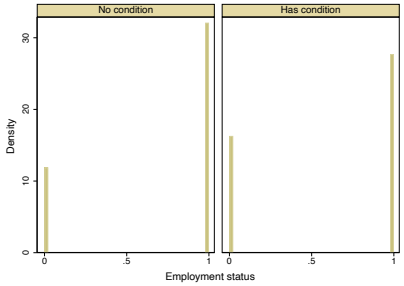


Figure 6.3.: Histograms of weekly hours worked for participants with and without cancer in Analysis A

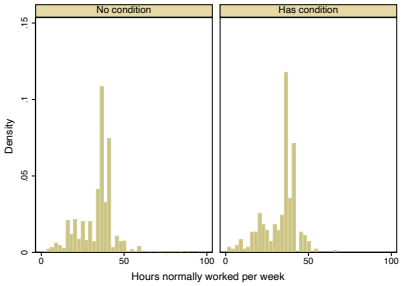
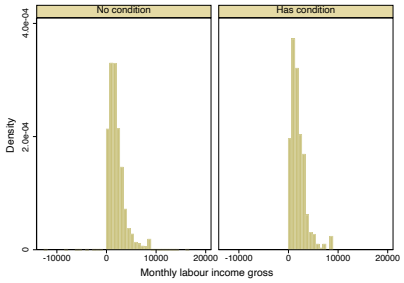
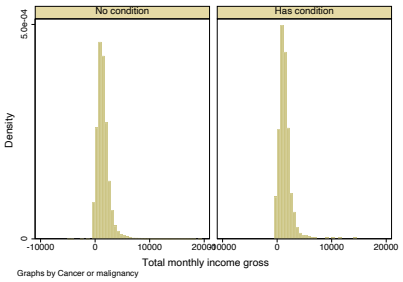


Figure 6.4.: Histograms of gross monthly earnings (£) for participants with and without cancer in Analysis A



Truncated at £20,000

Figure 6.5.: Histograms of gross monthly income (£) for participants with and without cancer in Analysis A



Truncated at £20,000

Figure 6.6.: Histograms of SF-12 (Short-Form 12) physical score for participants with and without cancer in Analysis A

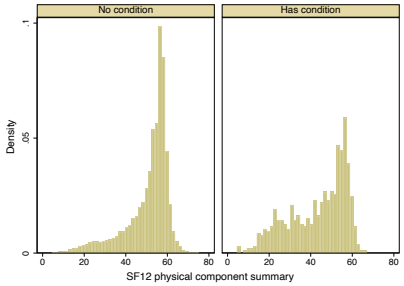
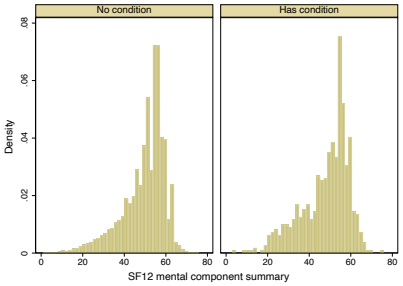
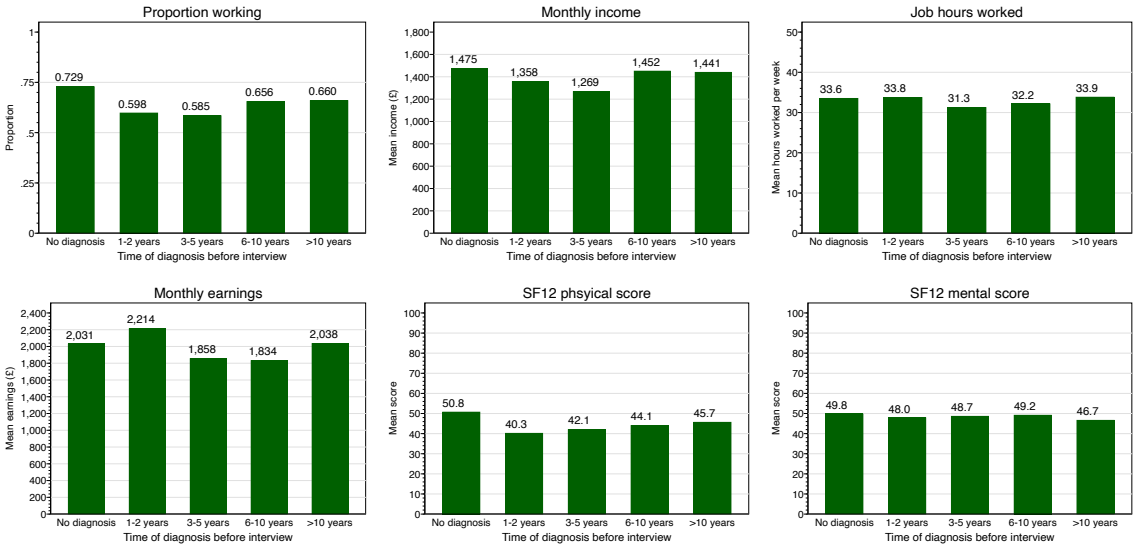


Figure 6.7.: Histograms of SF-12 (Short-Form 12) mental score for participants with and without cancer in Analysis A



Outcomes by time-from-diagnosis are compared to the no-cancer group in Figure 6.8. The first chart shows lower proportions in work for all exposed groups compared to the no-diagnosis group, with the lowest proportion observed in those who experienced a diagnosis 3–5 years before the interview. Income showed a similar pattern, with the lowest income observed in the 3–5 years group and lower incomes observed in all cancer groups than the no-diagnosis group. Hours worked for those in work showed little evidence of an association with cancer, with slightly higher hours worked in the 1–2 years and >10 years groups than the no-diagnosis group. There was little evidence of a notable drop in earnings for those in work, with higher earnings observed for the 1–2 years from diagnosis group and the >10 years from diagnosis group than the no-diagnosis group. SF-12 physical score declined by over 10 points between the 1–2 years group and the no-diagnosis group, with scores rising thereafter but remaining more than five points lower for the >10 years group. In SF-12 mental score, however, minimal differences between groups were observed.

Figure 6.8.: Outcomes stratified by time relative to diagnosis for participants with and without cancer in Analysis A



Note: SF12=Short-Form 12.

6.5.4 Analysis A: Regression Results for Analysis A

Results of univariable regression are shown in Table 6.4. Only for working and SF-12 physical scores were consistent significant associations with cancer observed. For working, the association had the largest magnitude at 3–5 years (OR 0.523; $p<0.001$), however the confidence intervals for the time periods showed considerable overlap,

suggesting cautious interpretation. For other outcomes, only monthly income at 3–5 years (β -£206.39; p=0.009) and SF-12 mental score at 10+ years (β -3.11; p<0.001) were statistically significant at the 5% level.

Multivariable models for working and SF-12 physical score are shown in tables 6.5–6.6. Adjustment for confounders decreased the magnitudes of associations between cancer and working, and their statistical significance, for all time groups with only the 1–2 years and 3–5 years groups retaining significant associations. While the association between cancer and employment remained when adjusted for comorbidities, the results showed larger and more significant associations for other health conditions. Adjusted for confounders, the SF-12 physical score had a significant negative association for all exposure times, but the association decreased with time from diagnosis. Significant associations were observed for SF-12 mental score at 1–2 years and at 10+ years from diagnosis. Among the outcomes job hours, earnings, log earnings the results were no more or less significant than in univariable models. Only log income at 1–2 years (β 0.18; p=0.001) was significant at the 5% level while monthly income was not significantly different from zero in any time category at the 5% level .

Table 6.4.: Summary of univariable and multivariable regression results for Analysis A

LOGISTIC MODELS		UNIVARIABLE			MULTIVARIABLE		
Outcome	**Explanatory Variable**	**Odds Ratio**	***p***	**AUC**	**Odds Ratio**	***p***	**AUC**
Working	No cancer diagnosis	1	Reference		1	Reference	
	Cancer at any time	0.631	<0.001	0.506	0.687	<0.001	0.697
	Cancer at t=1-2 years	0.553	<0.001	0.506	0.675	0.044	0.697
	Cancer at t=3-5 years	0.523	<0.001	0.506	0.571	0.001	0.697
	Cancer at t=6-10 years	0.707	0.041	0.506	0.721	0.052	0.697
	Cancer at t>10 years	0.721	0.025	0.506	0.780	0.112	0.697
OLS MODELS		**UNIVARIABLE**			**MULTIVARIABLE**		
Outcome	**Explanatory Variable**	**Coefficient**	***p***	**R^2**	**Coefficient**	***p***	**R^2**
Job hours	No cancer diagnosis	1	Reference		0	Reference	
	Cancer at any time	-0.711	0.202	<0.001	0.160	0.756	0.179
	Cancer at t=1-2 years	0.223	0.855	<0.001	0.645	0.563	0.179
	Cancer at t=3-5 years	-2.265	0.064	<0.001	-1.318	0.215	0.179
	Cancer at t=6-10 years	-1.362	0.218	<0.001	-0.471	0.642	0.179
	Cancer at t>10 years	0.298	0.749	<0.001	1.422	0.109	0.179
Monthly earnings (£)	No cancer diagnosis	1	Reference		0	Reference	
	Cancer at any time	-45.010	0.558	<0.001	-17.014	0.808	0.191
	Cancer at t=1-2 years	183.374	0.307	<0.001	263.283	0.086	0.191
	Cancer at t=3-5 years	-173.132	0.288	<0.001	-118.940	0.429	0.191
	Cancer at t=6-10 years	-197.193	0.156	<0.001	-245.209	0.054	0.191
	Cancer at t>10 years	7.147	0.958	<0.001	64.616	0.596	0.191
Log earnings	No cancer diagnosis	1	Reference		0	Reference	
	Cancer at any time	-0.012	0.832	<0.001	0.001	0.989	0.101
	Cancer at t=1-2 years	0.153	0.232	<0.001	0.197	0.096	0.102
	Cancer at t=3-5 years	-0.155	0.208	<0.001	-0.132	0.266	0.102
	Cancer at t=6-10 years	-0.080	0.465	<0.001	-0.107	0.318	0.102
	Cancer at t>10 years	0.033	0.714	<0.001	0.062	0.456	0.102
Monthly income (£)	No cancer diagnosis	1	Reference		0	Reference	
	Cancer at any time	-86.533	0.083	<0.001	9.715	0.881	0.098
	Cancer at t=1-2 years	-117.125	0.195	<0.001	70.812	0.493	0.098
	Cancer at t=3-5 years	-206.394	0.009	<0.001	-93.177	0.360	0.098
	Cancer at t=6-10 years	-23.086	0.840	<0.001	7.389	0.959	0.098
	Cancer at t>10 years	-33.913	0.731	<0.001	47.951	0.708	0.098
Log income	No cancer diagnosis	1	Reference		0	Reference	
	Cancer at any time	-0.113	0.273	<0.001	0.029	0.490	0.063
	Cancer at t=1-2 years	-0.317	0.248	<0.001	0.180	0.001	0.063
	Cancer at t= 3-5 years	-0.300	0.181	<0.001	-0.026	0.724	0.063
	Cancer at t=6-10 years	-0.046	0.818	<0.001	0.041	0.504	0.063
	Cancer at t>10 years	0.092	0.535	<0.001	-0.024	0.809	0.063
SF12 physical score	No cancer diagnosis	1	Reference		0	Reference	
	Cancer at any time	-7.384	<0.001	0.012	-6.133	<0.001	0.087
	Cancer at t=1-2 years	-10.532	<0.001	0.013	-8.840	<0.001	0.087
	Cancer at t=3-5 years	-8.733	<0.001	0.013	-7.270	<0.001	0.087
	Cancer at t=6-10 years	-6.659	<0.001	0.013	-5.547	<0.001	0.087
	Cancer at t>10 years	-5.059	<0.001	0.013	-4.075	<0.001	0.087
SF12 mental score	No cancer diagnosis	1	Reference		0	Reference	
	Cancer at any time	-1.815	<0.001	<0.001	-1.824	<0.001	0.049
	Cancer at t=1-2 years	-1.863	0.068	0.001	-2.022	0.036	0.049
	Cancer at t=3-5 years	-1.120	0.218	0.001	-1.097	0.224	0.049
	Cancer at t=6-10 years	-0.590	0.477	0.001	-0.756	0.335	0.049
	Cancer at t>10 years	-3.110	<0.001	0.001	-3.043	<0.001	0.049

Notes: AUC = area under (ROC) curve, SF12 = Short-Form 12, OLS = ordinary least squares.

Table 6.5.: Results of multivariable logistic regression on the association between cancer and the likelihood of working in Analysis A

Variable	**Odds Ratio**	**Std. Error**	***p***	**95% CI (lower/upper)**	
Cancer time from diagnosis					
No cancer diagnosis	1	Reference			
1-2 years	0.675	0.132	0.044	0.460	0.989
3-5 years	0.571	0.100	0.001	0.405	0.804
6-10 years	0.721	0.121	0.052	0.518	1.003
>10 years	0.780	0.122	0.112	0.574	1.060
Age in years	1.318	0.027	<0.001	1.265	1.372
Age^2	0.996	<0.001	<0.001	0.996	0.997
Age groups					
30-44 years old	1	Reference			
44-54 years old	1.528	0.094	<0.001	1.355	1.723
55-64 years old	1.648	0.183	<0.001	1.326	2.049
Sex					
Male	1	Reference			
Female	0.548	0.017	<0.001	0.516	0.582
Eduction level					
No higher qualification	1.000	Reference			
Higher level qualification	2.182	0.069	<0.001	2.051	2.322
Marital status					
Unmarried or separated	1	Reference			
Married or civil partner	1.567	0.047	<0.001	1.478	1.661
Ethnicity					
UK white	1	Reference			
Non UK white	0.455	0.015	<0.001	0.425	0.486
Parents working status aged 14					
At least one parent working	1	Reference			
No parents working	0.504	0.025	<0.001	0.457	0.556
Constant	0.021	0.009	<0.001	0.009	0.048
			area under ROC curve	=	0.697

Notes: CI = confidence interval, ROC = receiver operating characteristic, Std. Error = standard error.

Table 6.6.: Results of multivariable logistic regression on the association between cancer and SF-12 physical score in Analysis A

Variable	Coefficient	Std. Error	*p*	95% CI (lower/upper)	
Cancer time from diagnosis					
No cancer diagnosis	0	Reference			
1-2 years	-8.840	1.214	<0.001	-11.220	-6.459
3-5 years	-7.270	1.100	<0.001	-9.426	-5.113
6-10 years	-5.547	1.021	<0.001	-7.548	-3.545
>10 years	-4.075	0.863	<0.001	-5.766	-2.384
Age in years	-0.029	0.095	0.761	-0.216	0.158
Age^2	-0.002	0.001	0.043	-0.005	0.000
Age groups					
30-44 years old	0	Reference			
44-54 years old	0.154	0.250	0.539	-0.337	0.645
55-64 years old	-0.457	0.547	0.403	-1.528	0.615
Sex					
Male	0	Reference			
Female	-0.823	0.126	<0.001	-1.070	-0.576
Eduction level					
No higher qualification	0	Reference			
Higher level qualification	2.809	0.124	<0.001	2.566	3.052
Marital status					
Unmarried or separated	0	Reference			
Married or civil partner	1.263	0.134	<0.001	1.001	1.526
Ethnicity					
UK white	0	Reference			
Non UK white	-1.560	0.146	<0.001	-1.845	-1.274
Parents working status aged 14					
At least one parent working	0	Reference			
No parents working	-2.269	0.269	<0.001	-2.796	-1.742
Constant	56.211	1.931	<0.001	52.426	59.997
				R-squared = 0.087	

Notes: SF12 = Short-Form 12, OLS = ordinary least squares, Std. Error = standard error.

6.5.5 Analysis A: Additional Analysis

Table 6.7 shows further adjustment for comorbidities, while omitting the time categories to allow comparison of the relative contribution of cancer compared to other conditions. Although there was a significant association for cancer (OR 0.757; p=0.004), a larger reduction in the odds of working was observed in other conditions, for example, arthritis (OR 0.523; p<0.001), stroke (OR 0.344; p<0.001), epilepsy (OR 0.357; p<0.001), clinical depression (OR 0.331; p<0.001). Further results included comorbidities using the time groupings, but results did not differ substantially from

those without comorbidities. These results are shown in Appendix D.

Table 6.7.: Results of multivariable logistic regression on the association between cancer and working adjusted for comorbidities in Analysis A

Variable	Odds Ratio	Std. Error	*p*	95% CI (lower/upper)	
No condition	1	Reference			
Cancer	0.757	0.073	0.004	0.627	0.914
Asthma	0.895	0.048	0.038	0.805	0.994
Arthritis	0.523	0.027	<0.001	0.472	0.579
Congestive heart failure	0.427	0.124	0.004	0.242	0.756
Coronary heart failure	0.688	0.110	0.019	0.503	0.940
Angina	0.515	0.069	<0.001	0.397	0.669
Heart attack or myocardial infarction	0.580	0.082	<0.001	0.439	0.765
Stroke	0.344	0.051	<0.001	0.257	0.461
Emphysema	0.403	0.100	<0.001	0.248	0.655
Hyperthyroidism or an over-active thyroid	0.924	0.156	0.640	0.663	1.287
Hypothyroidism or an under-active thyroid	0.938	0.089	0.503	0.779	1.130
Chronic bronchitis	0.577	0.064	<0.001	0.465	0.717
Any kind of liver condition	0.586	0.070	<0.001	0.463	0.742
Diabetes	0.492	0.033	<0.001	0.431	0.561
Epilepsy	0.357	0.045	<0.001	0.279	0.458
High blood pressure	0.867	0.042	0.003	0.788	0.954
Clinical depression	0.331	0.019	<0.001	0.297	0.370
Other condition	0.933	0.049	0.181	0.842	1.033
Age in years	1.343	0.029	<0.001	1.287	1.401
Age^2	0.996	<0.001	<0.001	0.996	0.997
Age groups					
30-44 years old	1	Reference			
44-54 years old	1.552	0.099	<0.001	1.370	1.759
55-64 years old	1.732	0.202	<0.001	1.378	2.178
Sex					
Male	1	Reference			
Female	0.558	0.018	<0.001	0.524	0.595
Eduction level					
No higher qualification	1	Reference			
Higher level qualification	2.049	0.067	<0.001	1.922	2.185
Marital status					
Unmarried or separated	1	Reference			
Married or civil partner	1.423	0.044	<0.001	1.340	1.512
Ethnicity					
UK white	1	Reference			
Non UK white	0.424	0.015	<0.001	0.396	0.454
Parents working status aged 14					
At least one parent working	1	Reference			
No parents working	0.532	0.027	<0.001	0.481	0.589
Constant	0.017	0.008	<0.001	0.007	0.040
		area under ROC curve = 0.7375			

Notes: CI = confidence interval, ROC = receiver operating characteristic, Std. Error = standard error.

The results of the sensitivity analysis using sample weights are shown in Appendix D.2. When sample weights were incorporated into regression models the magnitude of the associations increased but not substantially. The significance of results increased, but

only for income at 3–5 years and job hours at 3–5 years did insignificant associations become significant at the 5% level.

6.6 Analysis B: Methods

6.6.1 Analysis B: Study Overview

Analysis B used a matched-pairs longitudinal study design, with difference in differences (DiD) estimation to estimate the average treatment effect on the treated (ATT). The study period for this analysis was 2008–18, corresponding to waves 1–9 of the UKHLS datasets. Exposure was a new cancer diagnosis in one of waves 4, 5 or 6. Other aspects of the setting were equivalent to those described in Section 6.4.1.

6.6.2 Analysis B: Participants

Analysis B: Eligibility Criteria

- Cancer cohort: a new cancer diagnosis during one of waves 4, 5 or 6 in UKHLS and no previous history of cancer.
- Control cohort: no history of cancer and no diagnosis in any of the nine waves.

Additionally, participants were required to be of working-age throughout the follow-up period: 18–64 years of age for men, 18–59 years of age for women. I used a lower age limit than in Analysis A because individuals with cancer were matched on age, and because the cancer diagnosis could only take place during the study period. Only individuals with a new (first) cancer diagnosed during the treatment period were selected for the treatment group, and no new cancers were included after the treatment period. All individuals with cancers diagnosed outside the treatment period were excluded.

Individuals with a new cancer in waves 4, 5 and 6 were matched with similar individuals. The controls were given a pseudo-event in each wave. Each individual was required to have at least one record prior to the event and at least one after the event. With nine waves in total, records for an individual could occur at a maximum of five waves before (t=-5) or after (t=5) the event. I matched non-cancer individuals to the cancer cohort by age, sex and wave of the study. Matching was carried out on an age variable that was precise to the nearest year, not the age categories described in Analysis A. As the number of individuals without a cancer diagnosis was relatively

large, it was possible to match exactly on age, sex and wave of the study. As in Analysis A, records obtained by proxy interview were excluded.

The size of the study was determined by the number of waves in which a new cancer diagnosis was measured. This exposure period was chosen to maximise the number of exposed individuals while retaining sufficient time in the follow-up and pre-exposure periods to assess trends in outcomes. While using the entire study period of nine waves would have achieved higher sample numbers of exposed patients, many would have few or no pre-diagnosis or post-diagnosis records with which to measure the ATT.

6.6.3 Analysis B: Variables

The outcome variables were the same as those in Analysis A. Because inflation affects monetary measures over time, I adjusted variables with monetary measures to 2009 levels to achieve comparability to the results of Analysis A. Other variables of Analysis A were used. Additionally, I used the following variables:

New cancer diagnosis during study period: Interviewees were asked *Since <last interview> has a doctor or other health professional newly diagnosed you as having any of the conditions listed on this card? If so, which ones?* To ensure that the cancer was a first cancer, I only included new cancers where the patient had no other previous record of cancer. The new condition variable was used to construct a binary variable with values: 0 = no new cancer diagnosis in this wave, 1 = new cancer diagnosis in this wave.

Treated: For a difference in differences analysis treatment and control groups must be defined. I defined my treatment group as individuals who reported a new cancer diagnosis in one of waves 4, 5 or 6, which was determined using the new health condition variable described above. A binary treatment variable was created with values: 0 = control (no cancer diagnosis during sample period or prior to it), 1 = treatment (where treatment was a cancer diagnosis at wave 4, 5 or 6).

Time after treatment event (t): The wave in which the cancer diagnosis occurred was assigned $t = 0$. In the control group, a pseudo-treatment event was also created with $t = 0$. The time of each longitudinal record in relation to the treatment event was then calculated. The t variable could take integer values in the range [-5, 5], however the treatment event record at $t = 0$ was removed as only post-treatment and pre-treatment values are required for DiD analysis. For regression analysis a dummy variable was created to show if a record belonged to the post-treatment period.

Analysis B: Quantitative Variables

The outcome and exposure variables were the same as those in Analysis A. I chose only to use unadjusted values for earnings and income, as the results of Analysis A showed no substantial difference in associations, while adjusted figures were more complicated to report and interpret without providing notable benefits.

6.6.4 Analysis B: Missing Data and Bias

Sample attrition has been noted in UKHLS, with only 52% of individuals participating for six years or more [265]. It was expected that the cancer cohort would suffer greater attrition due to higher mortality. In fact, the numbers of participants in subsequent waves showed greater attrition in the non-cancer cohort. This may have been due to high rates of attrition in young men, reported by another study [265]. Black people and those on lower incomes have also been reported as having had high attrition rates, while individuals with very poor health in wave 1 had higher attrition rates than others (though attrition rates varied little for other measures of self-assessed health) [265].

6.6.5 Analysis B: Statistical Methods

This analysis used DiD methods to estimate the ATT of employment outcomes for cancer patients. DiD methods are related to the potential outcomes framework [258] that deals with the problem of directly measuring causal effects from observational data: while we intuitively assign causal effects we can never observe them directly because counterfactual outcomes are missing. Taking $W = 1$ as the treatment unit, $W = 0$ as the control unit, $Y(1)$ the outcome value for the active unit, and $Y(0)$ the outcome value for the control unit, the causal effect = $Y(1) - Y(0)$ or some other relationship of $Y(1)$ to $Y(0)$. However, for a single unit W we only ever observe $Y(1)$ or $Y(0)$ and never both, making the evaluation of the treatment effect on a single unit impossible [258]. The potential outcomes framework sidesteps this difficulty by taking a statistical approach and observing outcomes over multiple units.

Extending the model, whether a particular individual i receives the intervention is given by D_i, Y_{it_0} is the outcome for i in the pre-treatment period and Y_{it_0} the outcome in the post-treatment period. The DiD estimator is equal to the ATT:

$$\begin{aligned} DiD = ATT = \{E(Y_{it_1}|D_i = 1, W_i = 1) - E(Y_{it_1}|D_i = 0, W_i = 0)\} \\ -\{E(Y_{it_0}|D_i = 1, W_i = 1) - E(Y_{it_0}|D_i = 0, W_i = 0)\} \end{aligned} \tag{6.6}$$

The ATT can be estimated using the regression model:

$$Y_i = \alpha + \lambda \mathrm{d}T_i + \gamma \mathrm{d}D_i + \delta \mathrm{d}T_i \cdot \mathrm{d}D_i + u_i \tag{6.7}$$

The ATT is given by δ, while Y is the employment outcome of interest, α is a constant, $\mathrm{d}T$ is a dummy for the treatment period, $\mathrm{d}D$ is a dummy representing whether an individual received treatment (which in this analysis is a cancer diagnosis), λ and γ are coefficients and u is a normally distributed random error term.

The DiD approach was appropriate for this study because there are clearly defined treatment and control groups, a clearly defined time in which the treatment occurs, and before and after periods. As a type of fixed-effects model, DiD allows for control of fixed unobservable characteristics, which makes it useful when time-invariant unobserved heterogeneity could confound a causal effect [267]. An alternative approach would be to use instrumental variables, but these are difficult to ascertain for cancer survivors, making a selection on observables approach more suitable [222].

The DiD model can be extended with matching to account for observed heterogeneity [267] though this can introduce bias if matching variables are correlated with the outcome variable [268]. In theory, confounders need not be controlled for in a DiD model unless they are associated with trends in the outcome, which follows from the parallel trends assumption. Confounding effects that are constant over time should be eliminated by differencing. For these reasons I matched on variables only likely to affect the trends in outcomes. Age would affect the trend in employment, as a 60 year old was more likely to retire within five years than a twenty year old. Sex was also likely to affect the trend, as the retirement age for women was five years lower than that for men during the study period. These factors are also associated with cancer. While comorbidities may affect both the likelihood of cancer and the likelihood of employment, it was not clear that they would affect employment trends. Individuals with health conditions may be more likely to leave employment but the effect on the trend in employment will be determined by age and thus already controlled for. Ideally, comorbidities would also be matched in waves before the cancer diagnosis to avoid

codetermination. The resulting low sample numbers combined with difficulties in matching prior to the diagnosis made this impractical. Propensity score matching is commonly used in DiD analysis, however, I lacked statistical knowledge in this area, and it was not clear to me how to interpret the diagnostic charts where most variables were categorical. Hence this method was not pursued.

For continuous outcomes, models were estimated using OLS. As estimates using matched pairs will produce clustered errors, standard errors that account for clustering were used. For binary outcomes with matched pairs, standard logistic models are inappropriate, and methods that account for matched samples are necessary. To account for matched pairs, I used conditional logistic regression to analyse binary outcomes. DiD estimation rests on the assumption of parallel trends, where those who received the treatment would, on average, have had the same outcomes as those who did not receive the treatment. This was assessed by visual inspection of graphical trajectories.

As a sensitivity analysis I carried out analyses using unmatched groups. This provided higher sample numbers than the matched analysis, and also minimised selection bias that may have occurred through matching. As individuals were not matched, there was no diagnosis-event for the control group, hence the treatment period was fixed within the study at waves 4–6, with pre-treatment and post-treatment periods also fixed. Compared to the matched analysis this meant a reduction in the lengths of the pre-treatment and post-treatment periods. Significance levels of 5% were used for all ATT estimates and 95% confidence intervals were reported. All analyses were carried out in Stata 15.1.

6.7 Analysis B: Results

6.7.1 Analysis B: Participants

Figure 6.9 illustrates how the study numbers were arrived at. Note that participants were required to be of working age for the entire study period, including the period before the diagnosis event in both the cancer and non-cancer groups.

Figure 6.9.: Flow chart of participant numbers for the cancer and non-cancer cohorts in Analysis B

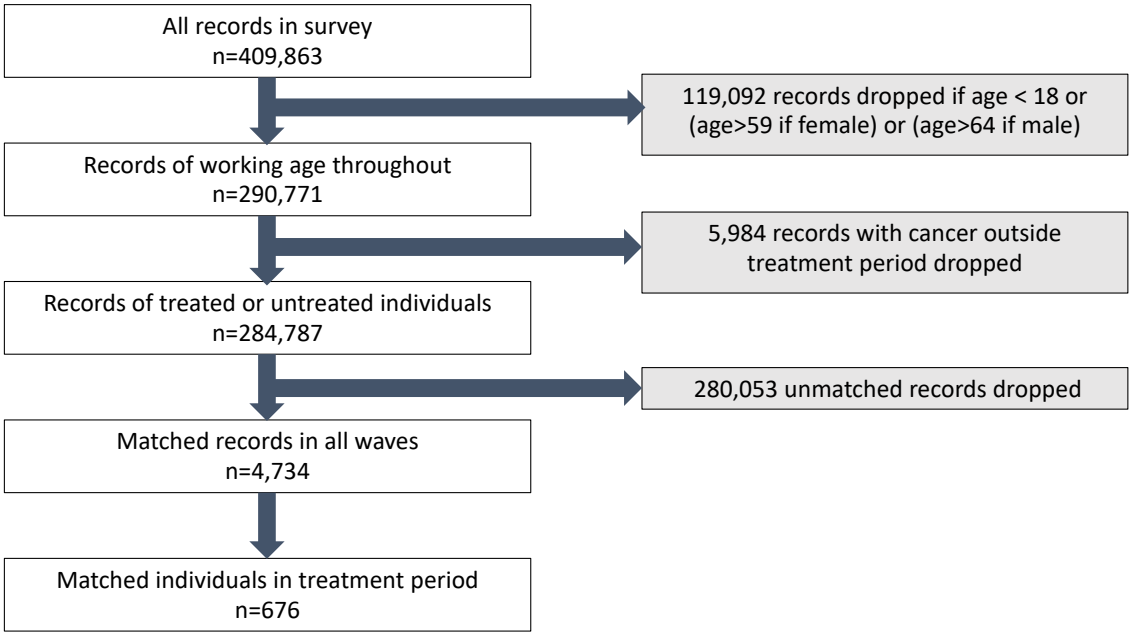


Table 6.8 shows the numbers of the matched exposed and control cohorts at each value of t. As the panel was not balanced, and because some patients had varying numbers of pre-event and post-event records, the numbers were not constant in each study period.

Table 6.8.: Numbers of records in each wave for the cancer and non-cancer cohorts in Analysis B

t	No cancer	Cancer	All records
-5	70	71	141
-4	176	182	358
-3	286	289	575
-2	305	319	624
-1	311	313	624
0	338	338	678
1	267	270	537
2	220	227	447
3	186	201	387
4	120	128	248
5	49	52	101
Total records	2,344	2,390	4,734
Total pre	1,148	1,174	2,322
Total post	842	878	1,720

Notes: pre = pre-diagnosis, post = post-diagnosis, t = time relative to diagnosis in years.

6.7.2 Analysis B: Descriptive Data

Characteristics of the participants for Analysis B are shown in Table 6.9. The characteristics were recorded during the interview for the wave of the diagnosis event, which was also the wave in which the participants were matched.

Table 6.9.: Sample characteristics measured at the time of diagnosis event for the cancer and non-cancer cohorts in Analysis B

	No cancer N=338	**Cancer** N=338
Age in years (mean, sd)	50.7 (8.3)	50.7 (8.3)
Sex		
Male	146 (43.2%)	146 (43.2%)
Female	192 (56.8%)	192 (56.8%)
Marital status		
Unmarried or separated	135 (40.8%)	146 (43.2%)
Married or civil partners	196 (59.2%)	192 (56.8%)
Employment status		
Not working	92 (27.4%)	119 (35.2%)
Working	244 (72.6%)	219 (64.8%)
Current economic activity		
self employed	34 (10.1%)	35 (10.4%)
Paid employment(ft/pt)	204 (60.4%)	173 (51.2%)
unemployed	19 (5.6%)	13 (3.8%)
retired	26 (7.7%)	27 (8.0%)
LT sick or disabled	24 (7.1%)	67 (19.8%)
doing something else	31 (9.2%)	23 (6.8%)
Ethnicity		
UK white	265 (79.6%)	296 (87.6%)
Non UK white	68 (20.4%)	42 (12.4%)
Higher level qualification		
No higher qualification	182 (56.0%)	188 (57.7%)
Higher level qualification	143 (44.0%)	138 (42.3%)
Hours normally worked per week (mean, sd)	34.0 (8.9)	32.8 (11.9)
Monthly labour income gross (mean, sd)	1611.7 (1823.5)	1290.7 (1571.8)
Total monthly income gross (mean, sd)	1782.0 (2211.0)	1573.9 (1554.8)
Health conditions count		
Zero conditions	189 (55.9%)	0 (0.0%)
One condition	75 (22.2%)	148 (43.8%)
Two or more conditions	74 (21.9%)	190 (56.2%)
SF12 physical component summary (mean, sd)	50.2 (11.1)	40.4 (13.5)
SF12 mental component summary (mean, sd)	49.8 (9.6)	45.8 (11.1)
No parents working at age 14		
At least one parent working	302 (94.7%)	304 (93.5%)
No parents working	17 (5.3%)	21 (6.5%)

Notes: sd = standard deviation, LT = long-term, ft = full time, pt = part time, SF12 = Short-Form 12. The separate health conditions listed in Analysis A are not shown due to changes in data coding after wave 1. Some variables do not sum to 338 due to missing values and are included for descriptive purposes but were not used in the analysis.

6.7.3 Analysis B: Outcome Data

Trajectories of Outcomes

Trajectories of outcomes in figures 6.10–6.15 showed considerable variation over time due to low numbers at each time point, making the assessment of parallel trends in the pre-treatment period challenging. There was evidence of an Ashenfelter dip—where the effect begins before the exposure—for employment. This is consistent with other results and may have been a reflection of prodromal symptoms [216] or a lack of reporting precision. All outcomes except hours worked showed dips at the treatment event time. While the SF-12 trajectories for the cancer group dipped at $t = 0$ then moved upwards towards the control group, employment trajectories continued to diverge throughout the follow-up.

Figure 6.10.: Trajectories of proportions working in the cancer and non-cancer cohorts in Analysis B

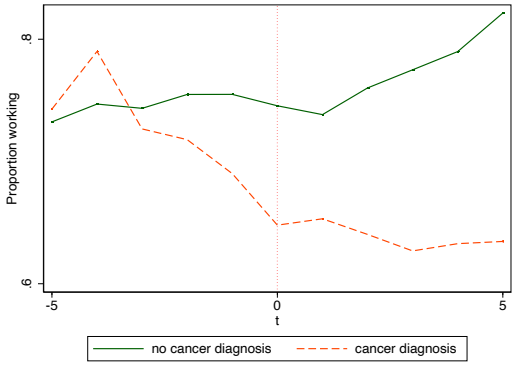


Figure 6.11.: Trajectories of weekly hours worked in the cancer and non-cancer cohorts in Analysis B

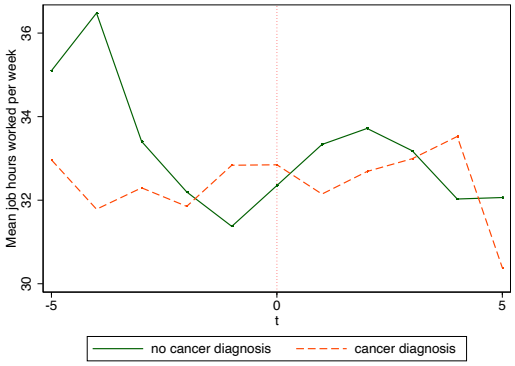


Figure 6.12.: Trajectories of monthly earnings in the cancer and non-cancer cohorts in Analysis B

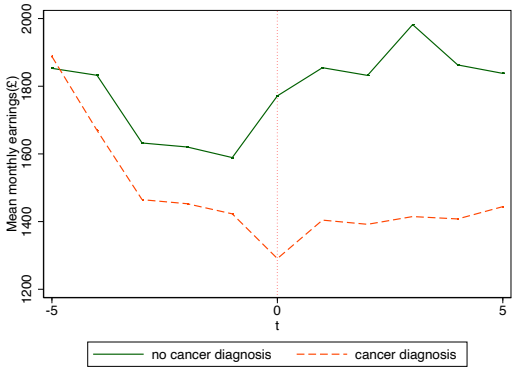


Figure 6.13.: Trajectories of monthly income in the cancer and non-cancer cohorts in Analysis B

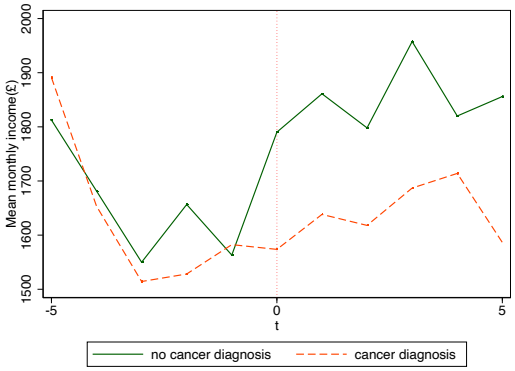


Figure 6.14.: Trajectories of SF-12 (Short-Form 12) physical score in the cancer and non-cancer cohorts in Analysis B

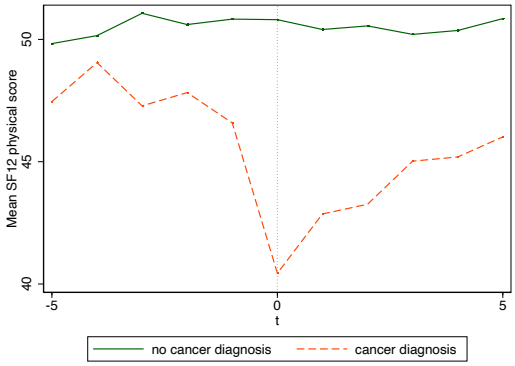
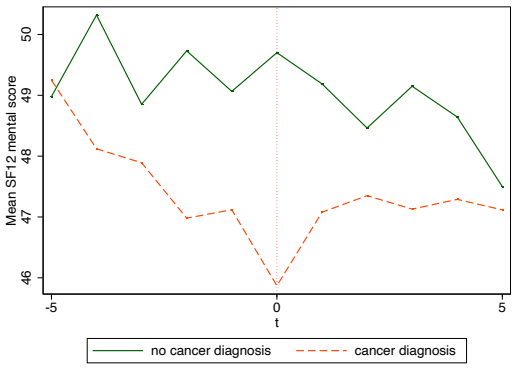


Figure 6.15.: Trajectories of SF-12 (Short-Form 12) mental score in the cancer and non-cancer cohorts in Analysis B



6.7.4 Analysis B: Difference in Differences Results

The results of the matched DiD analysis are shown in Table 6.10. ATT coefficients showed significant effects on working at 1–2 years and 3–5 years after diagnosis, with a larger association at 3–5 years. Earnings and income were also significantly affected while SF-12 physical showed a significant effect but SF-12 mental did not.

Table 6.10.: Results of difference in differences analyses on the average treatment effects on the treated in Analysis B

CONDITIONAL LOGISTIC						
Outcome		**ATT(OR)**	**Std. Error**	***p***	**95 % CI (lower/upper)**	
Working	all t>0	0.617	0.094	0.002	0.458	0.832
	0<t<3	0.651	0.100	0.005	0.482	0.879
	2<t<6	0.593	0.117	0.008	0.402	0.874
LINEAR (OLS)						
Outcome		**ATT**	**Std. Error**	***p***	**95 % CI (lower/upper)**	
Job hours	all t>0	0.222	0.960	0.818	-1.693	2.138
	0<t<3	0.041	1.006	0.967	-1.966	2.048
	2<t<6	0.457	1.235	0.713	-2.007	2.920
Monthly earnings (£)	all t>0	-316.965	124.281	0.013	-564.836	-69.094
	0<t<3	-240.875	107.844	0.029	-455.963	-25.787
	2<t<6	-414.381	173.921	0.020	-761.256	-67.506
Monthly income (£)	all t>0	-408.794	222.982	0.071	-853.517	35.930
	0<t<3	-544.607	368.866	0.144	-1280.286	191.073
	2<t<6	-234.185	136.075	0.090	-505.579	37.208
SF12 physical score	all t>0	-2.428	0.624	<0.001	-3.672	-1.183
	0<t<3	-3.550	0.630	<0.001	-4.806	-2.294
	2<t<6	-0.975	0.849	0.255	-2.669	0.719
SF12 mental score	all t>0	-0.074	0.738	0.920	-1.546	1.398
	0<t<3	-0.444	0.668	0.509	-1.776	0.889
	2<t<6	0.386	1.050	0.714	-1.709	2.482

Notes: ATT = average treatment effect on the treated, OR = odds ratio, OLS = ordinary least squares, SF12 = Short-Form 12, CI = confidence interval, Std. Error = standard error, t = time in years relative to diagnosis-event.

6.7.5 Analysis B: Sensitivity Analysis

Results of the analyses using unmatched groups are shown in Table 6.11. Due to the shorter post-diagnosis period, the results were not stratified by time and broadly concurred with the results of matched analysis although the ATT for earnings was not significant (ATT £59.418; p=0.545). Trajectories of outcomes are shown in Appendix E. Although the pre-diagnosis period was shorter, the greater smoothness of the trajectories due to larger sample numbers bolsters confidence in the parallel trends assumption.

Table 6.11.: Results of difference in differences analyses for unmatched samples on the average treatment effects on the treated in Analysis B

LOGISTIC				
Outcome	**ATT(OR)**	***p***	**95 % CI (low/high)**	
Working	0.632	<0.001	0.515	0.777
LINEAR				
Outcome	**ATT**	***p***	**95 % CI (low/high)**	
Job hours	0.940	0.203	-0.506	2.385
Monthly earnings (£)	59.418	0.545	-132.905	251.742
Monthly income (£)	-115.282	0.072	-241.071	10.507
SF12 physical score	-2.411	<0.001	-3.704	-1.119
SF12 mental score	0.255	0.658	-0.876	1.386

Note: ATT = average treatment effect on the treated, OR = odds ratio, OLS = ordinary least squares, SF12 = Short-Form 12, CI = confidence interval, Std. Error = standard error.

6.8 General Discussion

6.8.1 Key Results

This investigation measured the association between a cancer diagnosis and outcomes related to productivity, using both cross-sectional and longitudinal data to understand the dynamics of the outcomes over time. The results suggest that a cancer diagnosis lowers the likelihood of being employed and lowers self-reported physical health. The association with working was found to be significant up to five years after diagnosis in both analyses. A significant negative association between a cancer diagnosis and self-reported physical health was also measured up to 10 years after the diagnosis in Analysis A and up to five years after diagnosis in Analysis B. No consistent significant associations with job hours or self-reported mental health were found in either analysis, however, this again may have been due to low sample numbers rather than lack of effect. Significant negative associations with income and earnings were found in Analysis B, but in Analysis A were absent. The results of the two analyses broadly concurred in terms of direction and strength of the association for each outcome though differences in significance for some outcomes were noted. Interestingly, both analyses found the strongest negative association on the likelihood of working at 3–5 years after the diagnosis.

6.8.2 Interpretation

The results suggest that cancer can bring lasting declines in subjective physical health and a decrease in the likelihood of working, however, for those who remain in paid employment or return to it, the effect on hours worked may not be significant. The results do not suggest that cancer substantially reduces long-term earnings for those working, or income. Analysis A did not provide evidence of causality, as information on outcomes prior to the diagnosis was lacking, while the comparison of the results before and after the diagnosis in Analysis B provided evidence of a causal effect. Due to changing cohort characteristics and secular changes, the results from the two analyses could be expected to show variation, however the associations agreed in direction and magnitude to a notable extent.

Likelihood of Working and SF-12

My results suggesting declines in self-reported physical health and the likelihood of working were in line with other studies, which were summarised in Table 6.1. The results of the literature indicate that a cancer diagnosis leads to a reduction in general health [261, 228, 262], increased likelihood of unemployment [213, 219], increased likelihood of early retirement, and a lower probability of working [219, 216, 229]. The literature on health shocks also suggests that an exogenous shock to self-reported health [246, 269], such as that from a cancer diagnosis, leads to a lower likelihood of working. Of particular note in my results was that the strongest association with the likelihood of working was observed 3–5 years after diagnosis in both analyses. This may have been a statistical anomaly related to small sample sizes, or it could possibly have been related to the timing of government policy or social security benefits. A delayed effect on health is another explanation, but this seems unlikely. The occurrence of second cancers is an alternative possibility. Another explanation is that there was a cumulative effect of survivors leaving the labour market permanently. Cancer survivors tend to be older, consequently when they move out of the labour market they are less likely to return to it than younger individuals. This can lead to an increasing number being out of employment over time. Therefore, there would be no mean regression, in contrast to what was observed with self-reported health.

Other Outcomes

I did not find consistent significant associations with hours worked or with earnings, while other studies found these to be significantly reduced [270, 219, 216]. Zajacova et

al. (2015) [216] found that earnings for those in work dropped 40% within two years and remained low, while hours worked dropped 200 per year. The upper confidence interval extended to drops of only half this magnitude, but this is still a relatively large association and the discrepancy with my findings may be down to the characteristics of the samples. The differing characteristics of the US and UK populations may also limit external validity. This may also apply to Stone (2017) [270], who noted generalisability to other populations as a limitation. However, Candon (2015) [219] using English survey data, also found a larger effect of 4.2 fewer hours worked a week. This study had limitations similar to my own as it used self-reported survey data with a small sample size for the cancer cohort, and the significance of some results was under 5%. Furthermore, this study analysed older workers up to age 67, who may have been more likely to retire or move from full-time to part-time hours than younger ones. Alternatively, the discrepancy with my results may have been due to biases in UKHLS data leading to underestimated associations; for instance, the relatively high proportion of women. On the other hand, it may be that publication bias has led to the over-reporting of significant associations, while the true associations are less significant. If my results are accurate, a potential explanation is that female workers—who made up a higher proportion of the cancer groups than men—worked lower hours on average and had lower earnings. For workers who started with lower hours, it may have been more difficult to reduce hours further, e.g. by moving from full-time to part-time employment, meaning that these workers either had to maintain their hours or exit working entirely. Survivors may also have exited work entirely, either through early retirement or disability benefit, rather than reducing hours.

The high proportions of retired people and those who receive long-term sickness benefits in the cancer cohort support this possibility. I also found little evidence of an association between cancer diagnosis and declines in mental health. This was consistent with other findings for patients who remained cancer-free, but not for those with recurrence [228]. However, Helgeson's sample was mainly Caucasian and well-educated, whereas less-educated individuals may have lower financial resources and weaker coping mechanisms. Decreased mental health was reported elsewhere in the literature [213, 270, 261, 262]. Another possible source of discrepancy between my results and those of others, is that the majority of studies analysing work productivity in cancer patients analysed breast cancer patients [261], whereas my results aggregated a range of cancer sites affecting men and women. However, the high proportion of women with cancer in my sample, combined with the high prevalence of breast cancer for working-age women, suggests breast cancer may have had considerable weighting in my results.

6.8.3 Strengths

A strength of this study was the use of two distinct approaches to look both backward in time for evidence of association, and forward in time for evidence of a causal effect. These approaches allowed me to measure long-term associations with employment outcomes and HRQoL. An advantage of the UKHLS data was the large range of socioeconomic variables. This allowed for the adjustment of confounders such as ethnicity and education, that were not available in the linked NHS datasets in previous chapters.

6.8.4 Limitations

Sample Bias

This study also had limitations. UKHLS data have known biases, with young men, black people and the severely ill under-represented. Sensitivity analysis using sample weights suggested that the results may have underestimated associations, which could help to explain the disparities between my results and other studies that found larger associations. Weaknesses of survey data can include small relatively small sample sizes and response bias. Health data in UKHLS were limited, with information on the type and stage of cancer lacking. Although smoking variables existed in the data set, the high number of missing data led me to exclude them from the analysis. However, the data allowed me to adjust for a number of comorbidities and incorporate information on the temporality of health condition diagnoses. While self-reported health data are known to have biases, this tends to apply to measures of health with subjective scales, rather than more objective ones such as receiving a cancer diagnosis. Where subjective health measures were studied, the instruments used, such as SF-12, were designed to minimise such biases. Additional analyses to compensate for known biases in the data suggested that the results may underestimate associations, which may explain why expected reductions in hours worked and self-reported mental health were not observed.

Secular Factors

The study period came into proximity with the 2007–8 GFC. It is possible that cancer patients were more sensitive to the effects of financial crises, which could have influenced the observed associations, although the combination of cross-sectional and DiD analyses makes this less likely and strengthens the external validity of the results.

An additional limitation of survey data taken at yearly intervals, is that observation of employment at a particular time point is subject to survivorship bias. Most unemployment spells are short whereas observed spells tend to be long [271]. This could have caused overestimation of total employment, but whether that would lead to bias is unclear; a possible mechanism could be that cancer survivors move in and out of work more frequently. An additional source of survivor bias is related to deaths, as data are lacking on individuals who die during the survey, and those who survive may be statistically unrepresentative.

6.8.5 Changes to State Pension Age

Changes were made to the State Pension age (SPA) over the time of Analysis B. The SPA was 65 years of age for men and 60 for women until 2010 after which the age for women began to rise. The Pensions Act 2011 brought a new timetable for equalising the SPA for men and women, bringing the SPA for women to 65 by November 2018 [272]. It was not feasible to account for these changes in the analysis and my results excluded all women over 60. The exclusion of older working-age women means the results may underestimate the impact of cancer on employment for women.

6.8.6 Generalisability

While the data used a whole-of-UK sample with participants from England most highly represented, results should be relevant to Scotland which has the same social security system, although there are differences in the general health of the population, the health system, and the general employment rate. Comparison with other countries, particularly the US, may be more limited due to substantial differences in employment rates, social security systems and health systems. However, meta-analysis of studies covering Europe and North America has indicated that results may be comparable if age, diagnosis and background unemployment rate are fully controlled for [213]. These requirements were not fully satisfied in this study, suggesting that the results may better generalise to other high-income European countries.

6.8.7 Policy Implications and Future Research

This study suggested that a cancer diagnosis has long-term effects on employment for people in the UK. An interesting finding was the greater association with non-employment between three and five years after diagnosis. The reason for this result was not clear and is an area for further investigation.

Changes to State Pension Age

The age of state pension eligibility is rising in the UK and other developed nations, meaning more people of working age are likely to be living with the burden of cancer. Policymakers should consider the impact of increased cancer prevalence in people of working age as the SPA rises. At the least there should be additional support for cancer survivors who wish to continue working, while survivors unable to work should have the option of leaving the workforce with dignity and financial security.

Supporting Cancer Survivors

In this analysis cancer survivors who remained in work, or returned to it, tended not to reduce working hours. But survivors also had lowered self-reported health, suggesting that some continue working the same hours with an increased burden of ill health. Additional support may be required for survivors who wish or are compelled to maintain similar working hours as before their cancer diagnosis, as employment in favourable working conditions can aid recovery and reduce the risk of further illness [273]. Additionally, spells of unemployment can lead to a permanently lowered likelihood to work [254, 274] and to reduced health [274]. My results also indicated that incomes were not substantially affected. However, if work pensions become less prevalent, there may be an increased burden on social security if people leave work due to cancer while below the SPA.

Inequality

The increase in cancer incidence in working age people is likely to exacerbate inequalities. Manual workers may find it more difficult to continue working beyond 65 years of age than more highly paid knowledge workers. The high prevalence of female breast cancer could mean women are disproportionately affected. People with low socioeconomic status, who are at increased risk from cancer and chronic disease, may also have lower financial security, hence less able to take early retirement. Given these considerations, governments should give very careful consideration to the implications and viability of increasing the SPA, and should take measures to alleviate the burdens of increased cancer risk and its burdens in older workers.

Future Study

Table 6.2 shows a higher proportions of cancer survivors were in early retirement and on long-term sickness benefits than individuals without cancer in this sample.

However, the proportion of unemployed was lower in the cancer group, suggesting that lower overall employment in the cancer group was caused by transitions into early retirement and long-term sickness, which may have been a factor in the post-diagnosis employment trajectories. These factors were not statistically analysed in this analysis due to low numbers in sub-groups, however future studies could investigate transitions for cancer survivors between employment outcomes. Other areas for study include deeper analysis of the link between cancer, self-reported health and employment. The effects of changing ages for pension eligibility should also be examined as more data become available. Future research may also explore novel linkages such as between survey data and healthcare data.

6.8.8 Conclusion

In this study I used logistic regression, linear regression and difference in differences (DiD) estimation to measure long-term associations of cancer with working, self-reported health, and other outcomes related to employment. I found that cancer was associated with a lower likelihood of working up to five years after the diagnosis and with self-reported physical health at up to 10 years after the diagnosis. As cancer becomes more prevalent and the age of pension eligibility rises, more individuals will face the burden of cancer during their working lives. The implications of this will require the attention of policymakers for ensuring the well-being of working-age people with cancer. These results extend those of previous analyses in this thesis by highlighting the wider long-term costs of cancer. In the next and final chapter I will consider how the analyses described in chapters 3–6 met the objectives and overall aim of the thesis, and I will discuss the implications for policy and future research.

7 Overall Discussion

7.1 Introduction

The aim of this thesis was to enhance understanding of the long-term economic costs of cancer patients in Scotland using linked administrative data. The following research questions were asked.

1. How much healthcare do cancer patients in Scotland use and what are the associated monetary costs?
2. How do the costs vary over time?
3. Which factors influence costs and what is their relationship to survival??
4. How do costs compare to similar individuals without cancer?
5. What is the relationship between cancer and costs for patients with LTCs?
6. How does cancer affect other socioeconomic outcomes such as employment?

These questions gave rise to the following objectives.

1. Measure the healthcare use of cancer patients in Scotland and their associated costs.
2. Chart trajectories of costs over time.
3. Identify risk factors of costs and compare them to risk factors of survival.
4. Compare costs to patients without cancer and estimate excess costs.
5. Measure costs for patients with LTCs.
6. Measure the associations of cancer with wider socioeconomic outcomes such as employment.

In this chapter, I will summarise what I did to meet these objectives and answer the corresponding research questions. Results will be considered holistically in order to meet the overall aim of the thesis, and how to interpret them given overarching strengths and limitations. I will then discuss how far the results may apply in the future and to other populations, and consider the implications for policy of the results taken together with suggestions for future research.

7.2 Summary of Analyses

7.2.1 Chapter 3: Eight-Year Healthcare Resource Use of Cancer Patients in Scotland Using Linked NHS Datasets

To meet objectives 1–3, I linked NHS Scotland datasets SMR06, SMR00, SMR01 and PIS in order to measure the healthcare use of Scottish cancer patients over time. Eight-year incidence costs for each of the ten most common cancer types in Scotland and other cancers combined were estimated using per-episode costing with unit costs derived from the Scottish Costs Book. Trajectories of costs were charted by year and by phase-of-care. Risk factors for costs were determined using GLM regression and compared to risk factors of mortality derived from Cox regression.

I measured substantial costs over eight years for all cancer, with considerable variation by type. Mean eight-year costs varied considerably across cancer sites from £19,217 (95%CI £18,251 to £20,184) for malignant melanoma of skin patients to £47,672 (95%CI £45,500 to £49,843) for non-Hodgkin lymphoma patients. Trajectories of costs over time showed similar shapes across cancer types. At the cohort level, the most substantial proportion of costs accrued in the year after diagnosis. However, phase-of-care trajectories indicated that while cost-accrual rates were highest in the initial and end-of-life phases, higher overall costs were accrued in the continuing phase. Risk factors positively associated with higher costs tended to be negatively associated with higher survival but the relationship between costs and survival was not straightforward as the cancers with the highest costs tended to have moderate survival while the lowest costs and highest survival were observed in malignant melanoma of skin.

7.2.2 Chapter 4: Comparison with a Matched Control Group

To meet Objective 4 I measured the healthcare use and associated costs of patients with and without cancer, charted the trajectories over time, and calculated excess costs over eight years. I found that eight-year excess costs were positive for the 10 most common cancers and for all other cancers combined, ranging from £3,153 (95%CI £2,484 to £3,821) for patients with trachea, bronchus and lung cancers, to £34,106 (95%CI £31,623 to £36,589) for patients with non-Hodgkin lymphoma. Excess costs were positive in all phases of care for survivors who spent at least some time in a phase. However, across the entire cohort, excess costs turned negative at two years after diagnosis for cancers with high mortality rates (oesophagus and

trachea, bronchus and lung) while remaining positive during every year up to year 8 for less lethal cancers (breast, prostate, skin, head and neck, non-Hodgkin lymphoma), with other cancers turning negative between years 2 and 8.

7.2.3 Chapter 5: Inpatient Cost Trajectories For Patients With Long-Term Conditions

Further insights came from measuring inpatient use and associated costs for patients with underlying long-term conditions and charting them over time. This related to Objective 5. The associations of the diagnosis with costs specific to the LTCs were also measured and cost trajectories were charted. I found that cancer increased total inpatient costs for all LTCs and for a control group without any LTCs. The association with increased costs was of considerably greater magnitude (on a relative scale) for the comorbidity-free group with an increase of over 200% (cost ratio 3.016; 95%CI 2.975 to 3.058), than for any LTC, where increases ranged from 6.9% to 12.8%. The LTC-specific costs, however, were reduced after cancer in all LTC groups.

7.2.4 Chapter 6: Long-Term Employment Outcomes for Cancer Survivors in UKHLS

To gain a wider perspective on the costs to society and to meet Objective 6, I measured the long-term association between a cancer diagnosis and employment outcomes using logistic regression on cross-sectional survey data and difference in differences (DiD) analysis on longitudinal survey data. I found that the likelihood of working was significantly lower at the 5% significance level for cancer survivors than non-cancer survivors at 3–5 years after the diagnosis, with a 43% reduction in the odds of working after adjustment for confounders (OR 0.57; p=0.001). DiD analysis found a 41% reduction in the odds of working at 3–5 years after a cancer diagnosis (OR 0.59; p=0.008) for working-age individuals matched on age and sex. Beyond this, the likelihood was also lower, but the association was not statistically significant. Self-reported health was significantly lower up to 10 years beyond the diagnosis. However, the associations with other employment outcomes hours worked, earnings and total income were not statistically significant beyond one year after the diagnosis.

7.3 Overall Findings

The analyses provided evidence that the substantial economic costs found in other countries also occur in Scotland. At the cohort level most costs accrued in the year after diagnosis and diminished sharply thereafter, turning quickly negative for highly lethal cancers but remaining elevated for others up to eight years after diagnosis. There was considerable variation in costs across cancers which cannot be explained by survival alone. Both patient factors and cancer-related factors affected costs. The relationship with comorbidities was complex; while comorbidities can be expected to raise healthcare use, they also may reduce survival which can reduce long-term costs. The impact of cancer may be greater for younger, healthier people, particularly when the impact of employment is taken into account. The linkage of public-sector datasets can bring novel insights but the data can be challenging to access and may lack relevant variables. Survey data can be used to fill gaps in knowledge, however, these data also have limitations such as a lack of clinical variables. A limitation in my analyses was that I was unable to obtain employment data for the Scottish population. Hence my analysis of employment outcomes after a cancer diagnosis was carried out on a UK-wide sample, while healthcare outcomes were carried out on Scottish healthcare data.

Despite considerable heterogeneity, costs were substantial for all common cancers and other cancers combined. Even over eight years, cumulative costs were higher than those of a control group for all cancers, including those with very low survival rates. The highest total and excess costs were observed in non-Hodgkin lymphoma while the lowest total costs were observed in skin cancer and the lowest excess costs in trachea, bronchus and lung cancers. Excess costs were positive for all cancers over eight years, even for cancers with very low survival. Yearly excess costs turned negative for highly lethal cancers only two years after diagnosis, but the higher costs in year 1 outweighed the cumulative effects of negative costs during the eight-year follow-up. If the follow-up had been longer, it seems likely that overall excess costs would eventually have turned negative. This may have been even more likely if other healthcare such as mental health and social care had been incorporated, due to higher patient numbers in the non-cancer cohort in later years. Considered in terms of patient trajectories, cancer patients had higher rates of resource use in all phases of care than non-cancer patients, particularly the treatment and end-of-life phases. Costs were a result of the rate of resource use, the type of resource use, and the time spent in each phase (which may have been zero). While the average time spent in each phase was shorter than that of non-cancer patients, this was offset by higher rates of cost acquisition. The

time spent and rate of cost acquisition in each phase was determined by patient factors and characteristics of the cancer. These tended to be strongly linked to the cancer site, making it challenging to attribute costs to particular cancers.

7.4 Interpretation

7.4.1 Cost of Illness

A problem with comparing the costs of an illness across studies is the discrepancy in methods, research questions, outcomes, populations and health systems [80]. Even at the underlying theoretical and semantic levels, it is often unclear what is meant by *cost* in COI studies. Furthermore, it is seldom acknowledged that eliminating an illness would bring additional costs such as higher demand for social care and pensions. Opportunity costs are seldom included [77]. While this has justification in the fact that disease is a problem afflicting society rather than a good to be chosen against competing alternatives, it overlooks that a large proportion of the costs of a disease are the resources chosen to expend on treating it. Therefore, it should not be presumed that the more costly a disease, the more resources should be devoted to it. The cost to low-income and middle-income countries may also be underestimated because countries with limited healthcare will incur lower measured costs, while the relative economic impact may be greater due to lost output [275].

7.4.2 Unit Costing

I chose to use costs derived from the Scottish Costs Book rather than HRG costings from NHS England, as the per-episode method was more straightforward and it was not clear that English NHS costs would transpose to NHS Scotland. The Scottish Costs Book is a very broad collection of documents aimed more at hospitals than researchers, making the extraction of cost information challenging. Making the Scottish Costs Book more transparent to researchers would greatly aid the creation of costing studies using Scottish healthcare data.

7.4.3 Discounting

The reasons for and against discounting were described in Section 2.3.6. I chose to present undiscounted costs throughout, as these represented the actual costs incurred rather than the present value of future costs that discounted costs would represent. Discounting costs at higher rates will alter the cost magnitudes of different cancers

due to differential survival, and, for the same reason, will affect costs in excess of people without cancer. As everyone must eventually die of something, end-of-life costs will still occur without cancer but will be discounted due to occurring further in the future. Costs occurring over the long term, such as prescriptions, mental health and social care, will also be discounted further into the future. While individuals and businesses may discount future costs, it is not clear that government healthcare providers should do so if the goal is to anticipate future resource use. Doing so could underestimate the costs of long-term chronic disease compared to cancer, and the costs of cancers with high survival compared to more lethal ones.

7.4.4 Cost Trajectories

High resource use in the end-of-life phase is common for all patients but may be higher in cancer patients due to curative treatments that were unsuccessful, chemotherapy and other palliative treatment. Additionally, the higher mortality of cancer patients leads to greater numbers of end-of-life phases in the yearly trajectories. The highly elevated costs in the year after diagnosis will have included many end-of-life phases in addition to treatment phases, particularly in cancers with high one-year mortality. It is worth noting that the phases of care were, to some extent, defined arbitrarily and the lengths of phases were kept constant across cancers to simplify analysis and cross-cancer comparisons. For more detailed analysis of individual cancers, alternative phase lengths may be more appropriate .

7.4.5 Risk Factors

The relationship between stage and costs was non-linear, with stage III cancers associated with the highest costs. This may have been related to the patient trajectories; early stage cancers will often be treatable at relatively low cost, and will have higher survival than later stage cancers. Costs may be elevated in all phases compared to non-cancer patients, but fewer individuals will enter the costly end-of-life stage. Late-stage patients, by contrast, will be less likely to undergo curative treatment and more likely to die sooner. More patients will enter the costly end-of-life stage but the treatment and continuing phases will be shorter or never entered. Patients with stage III tumours will be more likely to undergo curative treatment, incurring high costs in the treatment phase. Those who survive through the continuing phase have elevated costs, while a considerable portion will enter the costly end-of-life phase. Patient factors and cancer characteristics will influence the type and amount of treatment given and the relative time spent in particular care phases, contributing to

variance in cost for individuals and heterogeneity across cancer sites. Treatment costs, such as drugs and radiology, will also vary by cancer site, causing additional heterogeneity, which may have been underestimated in this study due to the coarseness of the unit costs I used. Patients with lower-stage cancers were also more likely to be younger and suffer less chronic disease, which may have further lowered costs. An additional possibility is that the shock of a cancer diagnosis prompts positive lifestyle changes, as has been observed in some cancer patients [276], that would improve overall health and reduce healthcare use. It is also possible that unmeasured cancer and patient characteristics contributed to costs. Higher costs were generally associated with lower age and lower comorbidity, which may have been because cancers affecting young people were severe enough to have high costs, yet tended to be treated curatively due to lower stage and better patient health. The relatively high survival combined with higher resource use would lead to higher costs, with recurrence more likely due to longer survival, further increasing costs.

The highest total and excess costs were observed in non-Hodgkin lymphoma. High costs for this disease have been observed in other studies [179, 21]. Reasons may include the younger age of patients relative to other cancers, which contributes to greater likelihood of curative treatment [277]. Treatment failure in non-Hodgkin lymphoma has been observed as particularly costly [278]. Disease-specific clinical factors may also be contributors to high costs [279].

7.4.6 Data

Limitations of administration data were discussed in Section 2.4, the main drawbacks being a lack of socioeconomic variables and limited data on comorbidities. Survey data can provide socioeconomic variables that clinical data lack such as employment variables, but also suffer from a lack of data on comorbidities as well as tending to lack clinical data. UKHLS has some linkage between survey data and clinical data however, at the time of analysis the numbers linked to cancer were too low to be useful for analysis in this thesis. The use of administrative data also entailed challenges relating to information governance, such as ethics and legality, particularly when non-healthcare data such as Census and HMRC are sought. There were also technical challenges such as linking datasets external to the CHI system by probabilistic matching. Very serious consideration had to be given to the benefits of the research, and the application process was extensive and involved. The lengthy approval times for outcomes beyond healthcare prevented the use of these data within my PhD time frame. Furthermore, the uncertainty around what data could be

accessed caused considerable problems in planning the thesis and reviewing the literature, as the research questions were to some extent dependent on what data would ultimately be available to analyse. My initial proposal was based on being able to access anonymised public data on employment outcomes early in the project. When it became known that being granted access to such data would take considerable time and was far from certain, this necessitated readjustment of priorities. I sought out publicly available data on employment outcomes which led to the UKHLS study in Chapter 6. Providing a fuller account of disease costs requires access to many variables over a long time frame, making prospective studies costly. Administrative data can provide useful information on health outcomes but in the UK presently there are considerable challenges in combining wider socioeconomic outcomes which may make prospective study or longitudinal analysis using survey data more attractive.

7.4.7 Wider Economic Outcomes

Objective 6 of the study aimed to assess the broader economic impacts of employment. However, the absence of Scottish data posed a significant challenge in achieving this objective. Consequently, the analysis was limited to the available data from the UK. Despite this limitation, the study managed to explore the relationship between long-term employment and its wider economic implications in the UK. The absence of Scottish data undermined the comprehensiveness of the analysis as data from the UK may not fully capture the specific characteristics of the Scottish population, particularly health and socioeconomic ones. Consequently, the findings and conclusions drawn from this study should be interpreted with caution and may not be fully generalizable to the Scottish context.

Nonetheless, the analysis revealed a discernible link between cancer and long-term employment, with a notable result at 3–5 years after diagnosis that is worthy of further investigation. However, the magnitude of the association was not particularly high compared to other diseases. It is important to note that, as the state pension age continues to rise, the situation is expected to deteriorate further. This suggests that the future will hold more challenges associated with the long-term employment of people with cancer.

7.5 Strengths and Limitations

7.5.1 Strengths

A major strength of the analyses was the use of patient-level clinical data covering all of Scotland. All malignant cancers were included, with stratifications across the 10 most common cancer types and other cancers combined. This provided new information on the costs of less studied cancers, in addition to insights gained into cancer costs in the Scottish population and how costs change over time. The existence of a unique identifier, linked across NHS datasets, allowed the creation of a well-matched cohort across the whole Scottish population. The use of clinician-reported data is believed to reduce biases that affect self-reported data [269]. Regression methods were chosen to handle the non-normal distributions of healthcare costs [133], while accounting for clustering of errors due to loss of independence of observations in matched data. The inclusion of prescription data brought new insights into healthcare costs, as these data are not commonly included in other studies. Prescription costs tend to accrue by relatively small and numerous increments over time, thus improving statistical properties by avoiding the high mass of zero costs normally present in healthcare costs.

7.5.2 Limitations

Datasets

There were, however, limitations in the analyses. I had initially aimed to study employment outcomes for the Scottish population using linked administrative data. However, this turned out to be unachievable within the project time frame due to a combination of long application times and Covid-related access limitations. Hence I analysed employment outcomes using the publicly available UKHLS dataset. However, this comprised a UK-wide sample, with characteristics that may have diverged from the Scottish population. Additionally, these data were self-reported hence clinical variables were lacking, and therefore no breakdown by cancer site was possible. Administrative data are less likely to have such issues but have other limitations such as a lack of variables that describe socioeconomic status. Socioeconomic variables were limited in the SMR datasets, particularly compared to the UKHLS dataset, and those present, such as employment status, tended to have poor completion (although this was also true for many variables in UKHLS). Completion of SIMD was high, however SIMD is a postcode-level variable, used as a proxy for deprivation, and may not accurately

represent the deprived status of individuals [81]. Perhaps more surprising was the lack of some important clinical variables such as frailty. To some extent this may have been particular to my dataset, however, under-reporting of comorbidities has been observed in SMR data and is a noted issue with administrative data generally [118]. A further problem was the lack of SMR data for patients who arrived in Scotland before 2005 [81], which would lead to underestimations of comorbidity and outcomes.

Data Linkage

While Scotland's integrated health system makes linking datasets from distinct services possible, a desire for improvements in linkage has been stated [16]. The addition of data on primary care, mental health and social care would have provided a fuller account of healthcare use, however linking these datasets presented technical and ethical challenges that were considered insurmountable within the project time frame, and the additional costs may have outweighed the benefits provided. Including PIS data provided a fuller account of costs, particularly over the longer term, and improved distributional aspects of costs. But the inclusion came with drawbacks. Due to the very large number of records, the data were aggregated into annual records, which meant it was not possible to precisely locate costs on the patients' trajectories. Despite the aggregations, the number of records was still very high, which led to processing issues such as out-of-memory errors. Additionally, the PIS dataset only covered eight years. As pre-diagnosis costs were not an outcome under investigation, while a goal was to explore long-term cost trajectories, an eight-year follow-up period was chosen. However, differences in pre-diagnosis costs between exposure cohorts and cancer types showed interesting differences. With hindsight, a five-year follow-up with three years of pre-diagnosis costs that included prescriptions may have been more straightforward to calculate and report, although offering shorter trajectories. The benefit of the chosen approach was the insight into the longer-term outcomes that it provided.

Unit Costing

Although less precise than methods using HRGs, per-episode costing is believed to avoid biases associated with other costing methods [108]. While the unit costing method was likely to underestimate skew due to the lower variability of per-episode costs compared to per-diem methods, distributional accuracy was considered of lesser importance than unbiased means as these are more useful for cost estimates and evaluations [96]. Across an entire population, the estimated mean can be expected to asymptotically approach the true mean. However, this would hold only for all cancers

taken together, as the Scottish Costs Book did not break down cost units by cancer type. This means that differences in costs across cancer types were likely to be underestimated.

Secular Issues

The time frames of all analyses came into proximity with the GFC. This may have affected unemployment through a "social risk effect" and health through a "healthcare effect", with additional effects on healthcare costs [280]. The healthcare effect entails a rise in healthcare needs due to socioeconomic factors, while healthcare provision drops due to funding cuts. This would suggest the expenditure measured in this thesis may have underestimated healthcare needs. However, the healthcare effect is thought to have been relatively small at around 5% in the most affected countries: Greece, Ireland, Latvia and Portugal, and lower in the UK [280]. Measuring associations with employment rather than unemployment should have made results less partial to the social risk effect. Additionally, as austerity measures have not been reversed and major changes seem unlikely, the secular validity of results may be stronger than if they had been recorded prior to the GFC.

Sample Size Considerations

While the large sample sizes provided strong statistical power, this necessitated caution when interpreting results, as p-values and confidence intervals tended to be small. Therefore the magnitudes of associations took on greater importance compared to statistical significance. While claims have been made that large sample sizes may falsely report significant findings [281], others argue that large samples provide better protection against inflated effect sizes, improve overall statistical properties [282, 283] and allow analyses in important subgroups such as cancer site.

7.6 Generalisability

The literature shows considerable heterogeneity of costs across studies. This seems to be an inevitable result of variant populations, health systems, research questions and study methodologies. Patterns of results are more likely to generalise than a single monetary figure, which partly explains why this thesis put considerable emphasis on how costs vary across cancers and over time.

7.6.1 Population Factors

The health of Scotland's population is poorer than other western European countries [59, 16, 51], with social deprivation [57] and cultural differences [61] thought to be underlying factors for lower life expectancy and higher morbidity. Cancer mortality is also higher in Scotland than in England [60]. The impact of these differences on costs is uncertain; while higher morbidity requires more healthcare, this may be offset by lower longevity. Scotland's health in some respects bears a closer resemblance to the US than much of western Europe, with high levels of obesity and poor diet [56], however, comparability with the US is complicated by differences in the healthcare systems. The US lacks universal healthcare yet has substantially higher per-capita expenditure and relies on a large private insurance sector in addition to public healthcare. While many cancer drugs are developed and manufactured in the US, the cost to US payers tends to be higher as they are determined privately, whereas outside the US healthcare providers tend to have pricing power [44]. Hence the costs described here are likely to be lower than in the US, while units of resource use may be more comparable. High-income countries with universal (or near-universal) public healthcare, and with similarly sized populations include the other UK nations, Nordic nations, New Zealand and, to some extent, Ireland [284]. In terms of population health, England, Wales, Northern Ireland, New Zealand and Greece are closer to Scotland than the Nordic nations and Ireland. The remoteness of Scotland's rural areas may reduce comparability to England, while strengthening comparability with Wales, Ireland, the Nordic countries and New Zealand.

7.6.2 Other Factors

In addition to the challenges in making cross-country comparisons, secular factors complicate comparisons across time. In particular, the Covid pandemic has severely disrupted cancer care and other healthcare services. In England, the suspension of diagnostics and screening [285], combined with lowered primary care referrals [286] have been estimated to increase the stage of detection [285, 287], waiting lists [286] and deaths [285, 287]. If similar patterns occur in Scotland this is likely to affect costs, although the direction is uncertain. Covid has also altered employment patterns. While the proportion of economically inactive people aged 50–64 in the UK fell from over 38% in the mid-1990s to 25.5% in 2019, it has been rising since then to 27.7% in 2022 [288]. Other secular factors to consider are an ageing population, which will increase healthcare use, while reductions in smoking will increase longevity but with poorer health due to rising adiposity. New cancer treatments that prolong life rather

than prevent or cure could also raise costs.

Declining tobacco smoking may improve Scotland's health, making comparisons with other countries more relevant, however that could reduce the future validity of studies on past data. It is not clear what the societal costs and benefits of reduced smoking will be. Although we may expect a more productive workforce in better health, there may be an increased strain on healthcare and other services such as social care and pensions. Political factors such as changes to benefit systems and political shocks such as Brexit may also reduce generalisability of results over time. Although the comparability of overall costs across COI studies is low, other findings may be more comparable. General patterns of costs across cancers and over time are likely to generalise more than single results, while methods that adjust for comorbidities could aid comparisons of studies [179].

7.7 Policy Implications and Future Research

Cost Reduction

This thesis was concerned with measurement of health outcomes rather than evaluating interventions. Therefore the results are likely to be of more use in anticipating demand than reducing costs. However, the results do suggest areas where further exploration for cost savings could be made. For instance, the high costs of cancer for younger people in good health and how the costs for this group could be reduced are areas for further study. The high costs incurred in the continuing phase over the long term were also notable. Further research could focus on this phase specifically, investigating where costs are incurred and how they could be reduced.

Prevention

While it is tempting to say that the results in this thesis support efforts to focus on prevention, from a societal economic perspective this is not certain as costs saved in cancer treatment may simply accrue in other places like dementia care and pensions. However, if the impact on productivity is also accounted for, prevention becomes more attractive, particularly a broad approach focused on reducing common risk factors such as diet and physical activity. The Scottish Government's Cancer Action Plan 2023 to 2026 [289] has a welcome focus on these areas, in addition to measures on tobacco smoking reduction. However, the measures tend to be punitive price-based measures similar to the Scottish Government's minimum unit pricing (MUP) for alcohol. A

recent review of studies on MUP found that, while the policy reduced alcohol-related deaths, it exacerbated economic inequalities without significantly reducing hospital admissions [290]. Considering the strong links between the aforementioned risk factors and deprivation, a holistic approach to reduce deprivation should take priority over imposing additional hardship on people already struggling with high and inflating food costs.

Long-Term Care

The Scottish Government should also recognise that, as cancer treatment improves, there could be additional pressure on the NHS and other services due to the long-term healthcare needs of survivors. The proposed support for cancer victims tends to assume additional third sector provision. There needs to be a recognition that the improvement of cancer outcomes may require additional investment in healthcare, and support for cancer survivors in the workplace and in day-to-day life.

Age and Costs

While cancer is perceived as a disease of old age, the economic impact was found to be greater for younger people in my analyses. If the incidence of cancer in young people rises, and people are expected to retire at a higher age, the economic burden of cancer in working age people will increase. Reducing the economic costs of cancer may require a shift of focus towards younger people and people in good health.

Screening

Screening may reduce post-diagnosis costs, but should be targeted at detecting early-stage cancer, otherwise the added burdens of curative treatment and ongoing healthcare could lead to higher costs than for cancers with distant spread. There may be considerable potential for cost reductions in cancers beyond the four most common cancers, in particular non-Hodgkin lymphoma. Targeting early detection in young, healthy people may gain the greatest cost savings. Further investigation is required to evaluate the full costs and benefits of interventions.

Inequalities

A more challenging but possibly more effective strategy could be to tackle underlying socioeconomic factors such as deprivation, financial precarity and inequality. In addition to the effects of deprivation, inequalities in cancer outcomes could be

exacerbated by increases to the SPA. For instance, manual workers may find it more difficult to return to work after cancer treatment, while people with lower savings may lack the financial security to take time off work. Woman may be disproportionately affected due to the high prevalence of breast cancer in women of working age. Investigation of such possibilities will aid understanding of the effect of increasing longevity and government responses to it.

Data

Such investigations could be accelerated by an expansion of data-driven analysis through faster access to linked administrative data. However, the use of such data brings many challenges, and my experiences during this PhD suggest that the study of specific outcomes is more feasible than linking healthcare data to wider socioeconomic data. As NICE do not currently incorporate such data into economic evaluations, the benefits of synthesizing socioeconomic costs with healthcare costs are questionable, at least in the UK. In any case, researchers should be fully aware of the challenges involved if projects are to be planned and managed effectively, necessitating realistic estimates of access times based on previous applications with failure rates. The quality of Scottish NHS registry data could be enhanced by better reporting of stage and comorbidities. Additional variables on socioeconomic factors would be of benefit to researchers but may present difficulties in capturing accurately. While the Scottish Costs Book is a useful resource, it could be made more accessible to researchers unfamiliar with NHS administration; for example, explanations of terminology and better indexing of resources. It would also be helpful to have breakdowns on cancer types and more detail of the components of episode costs, as well as reporting of the uncertainty around costs.

7.8 Conclusion

This thesis used novel linkages of healthcare data to bring new insights into the long-term healthcare use and employment impact of cancer in Scotland. The large sample sizes allowed me to examine sub-groups, such as the 10 most common cancers in Scotland, and people with long-term conditions. The inclusion of a matched control group without cancer provided additional information on the economic burden of cancer and how it develops over time. Analysis of survey data provided insight into socioeconomic outcomes beyond healthcare, to give a broader account of the long-term costs of cancer. This knowledge will be valuable to policymakers,

economists, health professionals and other researchers in allocating healthcare, planning social policy, and understanding the long-term economic costs of cancer.

References

[1] C. Espina et al. "The essential role of prevention in reducing the cancer burden in Europe: A commentary from cancer prevention Europe". English. In: *Tumori*. 41st Oncology Days of the Organisation of European Cancer Institutes, OECI 2019. Italy. 105.2 Supplement (2019), pp. 54–57.

[2] J. Ferlay et al. "Cancer incidence and mortality patterns in Europe: Estimates for 40 countries and 25 major cancers in 2018". In: *European Journal of Cancer* 103 (Nov. 2018), pp. 356–387.

[3] Paul Hanly, Alison Pearce, and Linda Sharp. "The cost of premature cancer-related mortality: a review and assessment of the evidence". In: *Expert Review of Pharmacoeconomics & Outcomes Research* 14.3 (June 2014), pp. 355–377.

[4] Cancer Research UK. *Cancer Statistics for the UK*. 2021. URL: https://www.cancerresearchuk.org/health-professional/cancer-statistics-for-the-uk/worldwide-cancer (visited on 02/16/2021).

[5] John LaMattina. *Is Biopharma Investing Too Much In Cancer R&D?* 2018. URL: https://www.forbes.com/sites/johnlamattina/2018/04/24/is-biopharma-investing-too-much-in-cancer-rd/ (visited on 02/16/2021).

[6] Nick Bosanquet and Karol Sikora. "The economics of cancer care in the UK". In: *The Lancet Oncology* 5.9 (Sept. 2004), pp. 568–574.

[7] Carin A. Uyl-de Groot, Saskia de Groot, and Adri Steenhoek. "The economics of improved cancer survival rates: better outcomes, higher costs". In: *Expert Review of Pharmacoeconomics & Outcomes Research* 10.3 (June 2010), pp. 283–292.

[8] Caroline Anger et al. *Global Oncology Trends 2019*. Tech. rep. 2019. URL: https://www.iqvia.com/insights/the-iqvia-institute/reports/global-oncology-trends-2019.

[9] Bengt Jönsson et al. "The cost and burden of cancer in the European Union 1995–2014". In: *European Journal of Cancer* 66 (Oct. 2016), pp. 162–170.

[10] Jacob J.E. Koopman et al. "An emerging epidemic of noncommunicable diseases in developing populations due to a triple evolutionary mismatch". In: *American Journal of Tropical Medicine and Hygiene* 94.6 (June 2016), pp. 1189–1192.

[11] Public Health Scotland. *Cancer in Scotland*. Tech. rep. 2020. URL: https://www.isdscotland.org/health-topics/cancer/cancer-statistics/Cancer-in-Scotland-July-2020.pdf.

[12] D I Conway et al. "Socioeconomic factors associated with risk of upper aerodigestive tract cancer in Europe." In: *European journal of cancer (Oxford, England : 1990)*. [Comment in: Evid Based Dent. 2011;12(3):87-8; PMID: 21979774 [https://www.ncbi.nlm.nih.gov/pubmed/21979774]] 46.3 (2010), pp. 588–598.

[13] The World Bank. *Health | Data*. 2021. URL: https://data.worldbank.org/topic/health?locations=XD (visited on 04/22/2021).

[14] Office for National Statistics. *Living longer*. 2018. URL: https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/ageing/articles/livinglonger/implicationsofchildlessnessamongtomorrowsolderpopulation (visited on 02/18/2021).

[15] Office for National Statistics. *Subnational population projections for England*. 2018. URL: https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationprojections/bulletins/subnationalpopulationprojectionsforengland/2018based (visited on 02/18/2021).

[16] Organisation for Economic Co-operation and Development. *OECD Reviews of Health Care Quality: United Kingdom 2016*. Tech. rep. Feb. 2016. URL: https://www.oecd-ilibrary.org/social-issues-migration-health/oecd-reviews-of-health-care-quality-united-kingdom-2016%7B%5C_%7D9789264239487-en.

[17] National Records of Scotland. *Population Projections for Scottish Areas 2018-based*. Tech. rep. 2018. URL: www.nrscotland.gov.uk.

[18] Office for National Statistics. *Health state life expectancies*. 2018. URL: https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/healthandlifeexpectancies/bulletins/healthstatelifeexpectanciesuk/2016to2018 (visited on 02/18/2021).

[19] Office for National Statistics. *Adult smoking habits in the UK*. 2018. URL: https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/healthandlifeexpectancies/bulletins/adultsmokinghabitsingreatbritain/2019 (visited on 02/18/2021).

[20] Garth Reid et al. *Review of 'Creating a tobacco-free generation: A Tobacco Control Strategy for Scotland'*. Tech. rep. Public Health Evidence Network (PHEN), 2017.

[21] K. Robin Yabroff et al. "Cost of Care for Elderly Cancer Patients in the United States". In: *JNCI: Journal of the National Cancer Institute* 100.9 (May 2008), pp. 630–641.

[22] NHS National Services Scotland. *Cancer Mortality in Scotland*. Tech. rep. 2019. URL: https://www.statisticsauthority.gov.uk/national-statistician/types-of-official-statistics/.

[23] Nicholas Bosanquet and Karol Sikora. *The Economics of Cancer Care*. Cambridge University Press, 2006.

[24] Public Health Scotland. *Cancer Incidence in Scotland 28 April 2020*. Tech. rep. 2020. URL: https://beta.isdscotland.org/find-publications-and-data/conditions-and-diseases/cancer/cancer-incidence-in-scotland/28-april-2020/.

[25] Heather O. Dickinson. "Cancer trends in England and Wales: Good data and analysis are vital to improving survival". In: *BMJ* 320.7239 (Apr. 2000), pp. 884–885.

[26] Frederick K. Ho et al. "Changes over 15 years in the contribution of adiposity and smoking to deaths in England and Scotland". In: *BMC Public Health* 21.1 (Dec. 2021), pp. 1–8.

[27] NHS Digital. *Obesity-related hospital admissions*. 2021. URL: https://digital.nhs.uk/data-and-information/publications/statistical/statistics-on-obesity-physical-activity-and-diet/england-2021/part-1-obesity-related-hospital-admissions (visited on 05/19/2021).

[28] Office for National Statistics. *Adult smoking habits in the UK*. 2019. URL: https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/healthandlifeexpectancies/bulletins/adultsmokinghabitsingreatbritain/2019 (visited on 06/11/2021).

[29] Peter Scarborough et al. "The economic burden of ill health due to diet, physical inactivity, smoking,

alcohol and obesity in the UK: an update to 2006–07 NHS costs".
In: *Journal of Public Health* 33.4 (Dec. 2011), pp. 527–535.

[30] Food Standards Agency. *National Diet and Nutrition Survey*. Tech. rep. 2018.

[31] Karen L Barton et al. *Estimation of Food and Nutrient Intakes From Food Purchase Data in Scotland*. Tech. rep. 2018.

[32] Scottish Government. *The Scottish Health Survey 2021*. Tech. rep. 2021. URL: https://www.gov.scot/publications/scottish-health-survey-2021-volume-1-main-report/documents/.

[33] Rosalind A. Breslow and Kenneth J. Mukamal.
"Measuring the burden-current and future research trends: Results from the NIAAA expert panel on alcohol and chronic disease epidemiology".
In: *Alcohol Research: Current Reviews* 35.2 (2013), pp. 250–259.

[34] NHS Health Scotland. *Hospital admissions, deaths and overall burden of disease attributable to alcohol consumption in Scotland*. Tech. rep.
NHS Health Scotland, 2018. URL: https://www.scotpho.org.uk/media/1597/scotpho180201-bod-alcohol-scotland.pdf.

[35] Public Health Scotland. *MESAS monitoring report*. Tech. rep. 2021.
URL: https://www.publichealthscotland.scot/publications/mesas-monitoring-report-2021/.

[36] Office for National Statistics. *Adult drinking habits in Great Britain*.
URL: https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/drugusealcoholandsmoking/bulletins/opinionsandlifestylesurveyadultdrinkinghabitsingreatbritain/latest (visited on 07/28/2023).

[37] Office for National Statistics. *Cancer registration statistics, England*. 2021.
URL: https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/bulletins/cancerregistrationstatisticsengland/final2016 (visited on 02/19/2021).

[38] Melina Arnold et al.
"Global patterns and trends in colorectal cancer incidence and mortality".
In: *Gut* 66.4 (Apr. 2017), pp. 683–691.

[39] National Cancer Institute. *Cancer Trends Progress Report*. 2021.
URL: https://progressreport.cancer.gov/ (visited on 02/17/2021).

[40] National Cancer Institute. *SEER Explorer*. 2021.
URL: https://seer.cancer.gov/explorer/application.html (visited on 05/01/2021).

[41] Arnold L Potosky, Eric J Feuer, and David L Levin. "Impact of Screening on Incidence and Mortality of Prostate Cancer in the United States". In: *Epidemiologic Reviews* 23.1 (2001).

[42] B. Joseph Elmunzer et al. "Effect of Flexible Sigmoidoscopy-Based Screening on Incidence and Mortality of Colorectal Cancer: A Systematic Review and Meta-Analysis of Randomized Controlled Trials". In: *PLOS Medicine* 9.12 (Dec. 2012), e1001352.

[43] R.C.G. Russell and T. Treasure. "Counting the cost of cancer surgery for advanced and metastatic disease". In: *British Journal of Surgery* (2012).

[44] Kerstin N. Vokinger et al. "Prices and clinical benefit of cancer drugs in the USA and Europe: a cost–benefit analysis". In: *The Lancet Oncology* 21.5 (May 2020), pp. 664–670.

[45] Heinz Ludwig and Mangesh Thorat. "Optimum cancer care - An unaffordable goal?" In: *Lancet Oncology* 5.9 (Sept. 2004), pp. 529–530.

[46] R. Luengo-Fernandez, J. Leal, and A. M. Gray. "UK research expenditure on dementia, heart disease, stroke and cancer: Are levels of spending related to disease burden?" In: *European Journal of Neurology* (2012).

[47] Ramon Luengo-Fernandez, Jose Leal, and Alastair Gray. *Research spend in the UK*. Tech. rep. 2019. URL: https://www.stroke.org.uk/sites/default/files/sa-research%7B%5C_%7Dspend%7B%5C_%7Din%7B%5C_%7Dthe%7B%5C_%7Duk%7B%5C_%7Djuly2016%7B%5C_%7Dweb.pdf.

[48] Andrew G. Renehan and Anthony Howell. "Preventing cancer, cardiovascular disease, and diabetes". In: *Lancet* 365.9469 (Apr. 2005), pp. 1449–1451.

[49] Norman E. Sharpless. "COVID-19 and cancer". In: *Science* 368.6497 (June 2020), p. 1290.

[50] The World Bank. *Health Nutrition and Population Statistics | DataBank*. 2021. URL: https://data.worldbank.org/indicator/SH.DYN.NCOM.ZS?locations=GB-XD (visited on 04/22/2021).

[51] Jon Minton et al. "Visualising and quantifying 'excess deaths' in Scotland compared with the rest of the UK and the rest of Western Europe". In: *Journal Of Epidemiology And Community Health* 71.5 (2017), pp. 461–467.

[52] Tadeusz Dyba et al. "The European cancer burden in 2020: Incidence and mortality estimates for 40 countries and 25 major cancers".
In: *European Journal of Cancer* 157 (Nov. 2021), pp. 308–347.

[53] ISD Scotland. *Cancer in Scotland*. 2017. URL: http://www.isdscotland.org/Health-Topics/Cancer/Publications/2017-10-31/Cancer%7B%5C_%7Din%7B%5C_%7DScotland%7B%5C_%7Dsummary%7B%5C_%7Dm.pdf.

[54] Ramon Luengo-Fernandez et al. "Economic burden of cancer across the European Union: a population-based cost analysis."
In: *The Lancet. Oncology* 14.12 (2013), pp. 1165–1174.

[55] Paul McCrone. "Capturing the Costs of End-of-Life Care: Comparisons of Multiple Sclerosis, Parkinson's Disease, and Dementia". English.
In: *Journal of Pain and Symptom Management* 38.1 (July 2009), pp. 62–67.

[56] Frank Popham. "Is there a "Scottish effect" for self reports of health? Individual level analysis of the 2001 UK census".
In: *BMC Public Health* 6.1 (July 2006), pp. 1–11.

[57] G. McCartney et al. "Why the Scots die younger: Synthesizing the evidence".
In: *Public Health* 126.6 (June 2012), pp. 459–470.

[58] Linsay Gray and Alastair H. Leyland. "A multilevel analysis of diet and socio-economic status in Scotland: Investigating the 'Glasgow effect'".
In: *Public Health Nutrition* 12.9 (Sept. 2009), pp. 1351–1358.

[59] Nicola Jane Shelton. "Regional risk factors for health inequalities in Scotland and England and the "Scottish effect"".
In: *Social Science and Medicine* 69.5 (Sept. 2009), pp. 761–767.

[60] Gerry McCartney et al. "Explaining the excess mortality in Scotland compared with England: Pooling of 18 cohort studies". In: *Journal of Epidemiology and Community Health* 69.1 (Jan. 2015), pp. 20–27.

[61] R. O'Brien, K. Hunt, and G. Hart.
'The average Scottish man has a cigarette hanging out of his mouth, lying there with a portion of chips': Prospects for change in Scottish men's constructions of masculinity and their health-related beliefs and behaviours.
Sept. 2009. (Visited on 06/01/2021).

[62] Office for National Statistics. *Healthcare expenditure, UK Health Accounts*. 2018.
URL: https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/healthcaresystem/bulletins/ukhealthaccounts/2018 (visited on 02/18/2021).

[63] P. S. Hall et al. "Costs of cancer care for use in economic evaluation: a UK analysis of patient-level routine health system data".
In: *British Journal of Cancer* 112.5 (Jan. 2015), pp. 948–956.

[64] Martyn P. T. Kennedy, Peter S. Hall, and Matthew E. J. Callister. "Factors affecting hospital costs in lung cancer patients in the United Kingdom."
In: *Lung cancer (Amsterdam, Netherlands)* 97 (2016), pp. 8–14.

[65] Ian Kunkler. "Cure, palliation, and cost in cancer care".
In: *Lancet Oncology* 5.12 (Dec. 2004), p. 709.

[66] Julia M. Langton et al. "Retrospective studies of end-of-life resource utilization and costs in cancer care using health administrative data: A systematic review".
In: *Palliative Medicine* 28.10 (Dec. 2014), pp. 1167–1196.

[67] J. Huang et al. "Time Spent in Hospital in the Last Six Months of Life in Patients Who Died of Cancer in Ontario".
In: *Journal of Clinical Oncology* 20.6 (Mar. 2002), pp. 1584–1592.

[68] Cathy J. Bradley et al.
"Productivity costs of cancer mortality in the United States: 2000-2020."
In: *Journal of the National Cancer Institute* 100.24 (2008), pp. 1763–1770.

[69] Soheila Khorasani et al. "Years of potential life lost and productivity costs due to premature cancer-related mortality in Iran." In: *Asian Pacific journal of cancer prevention : APJCP* 16.5 (2015), pp. 1845–1850.

[70] Tracy Hampton. "Cancer Treatment's Trade-off".
In: *JAMA* 294.2 (July 2005), p. 167.

[71] Agnes Dumas et al. "The right to be forgotten: a change in access to insurance and loans after childhood cancer?." In: *Journal of cancer survivorship : research and practice* 11.4 (2017), pp. 431–437.

[72] Cheryl K. Altice et al. "Financial hardships experienced by cancer survivors: A systematic review". English.
In: *Journal of the National Cancer Institute* 109.2 (2017).

[73] Yakir Rottenberg and Angela G. E. M. de Boer.
"Risk for unemployment at 10 years following cancer diagnosis among very long-term survivors: a population based study".
In: *Journal of Cancer Survivorship* (Feb. 2020).

[74] Alan Shiell, Karen Gerard, and Cam Donaldson.
"Cost of illness studies: An aid to decision-making?"
In: *Health Policy* 8.3 (Dec. 1987), pp. 317–323.

[75] Lorna Guinness. "Counting the Costs". In: *Introduction to Health Economics*. Ed. by Lorna Guinness and Virginia Wiseman. Dawson Era, 2011, pp. 201–217.

[76] Thomas A. Hodgson and Mark R. Meiners.
"Cost-of-Illness Methodology: A Guide to Current Practices and Procedures".
In: *The Milbank Memorial Fund Quarterly. Health and Society* 60.3 (1982), p. 429.

[77] B.S. Bloom et al.
"Usefulness of US cost-of-illness studies in healthcare decision making".
In: *PharmacoEconomics* 19.2 (Sept. 2001), pp. 207–213.

[78] D. Chisholm et al.
"Economic impact of disease and injury: counting what matters".
In: *British Medical Journal* 340.mar02 1 (Mar. 2010), pp. c924–c924.

[79] Eberechukwu Onukwugha et al.
"Cost-of-Illness Studies: An Updated Review of Current Methods".
In: *PharmacoEconomics* 34.1 (Sept. 2015), pp. 43–58.

[80] Allison Larg and John R. Moss. "Cost-of-Illness Studies".
In: *PharmacoEconomics* 29.8 (Aug. 2011), pp. 653–671.

[81] Nazir Lone. "An evaluation of five year survival and major health care resource use following admission to Scottish intensive care units".
PhD thesis. Edinburgh University, 2013.

[82] Inuit Tapiriit Kanatimi. *Inuit & Cancer: Fact Sheets*. Tech. rep. 2009.
URL: https://www.itk.ca/sites/default/files/private/factsheet-seriesFINAL2.pdf.

[83] Ebere Akobundu et al.
"Cost-of-illness studies: a review of current methods". English.
In: *PharmacoEconomics* 24.9 (Sept. 2006).

[84] National Institute for Health and Care Excellence.
Guide to the methods of technology appraisal 2013. 2013. URL: https://www.nice.org.uk/process/pmg9/chapter/the-reference-case.

[85] Soeren Mattke et al.
"A Review of Methods to Measure Health-related Productivity Loss".
In: *AJMC* 13.4 (2007).

[86] Wei Zhang, Nick Bansback, and Aslam H. Anis.
"Measuring and valuing productivity loss due to poor health: A critical review".
In: *Social Science & Medicine* 72.2 (Jan. 2011), pp. 185–192.

[87] Bengt Liljas. "How to Calculate Indirect Costs in Economic Evaluations".
In: *PharmacoEconomics* 13.1 (1998), pp. 1–7.

[88] Fred Moseley. "Piketty and Marginal Productivity Theory".
In: *International Journal of Political Economy* 44.2 (Apr. 2015), pp. 105–120.

[89] Changik Jo. "Cost-of-illness studies: concepts, scopes, and methods". In: *Clinical and Molecular Hepatology* 20.4 (2014), p. 327.

[90] Marc A. Koopmanschap and B. Martin van Ineveld. "Towards a new approach for estimating indirect costs of disease". In: *Social Science & Medicine* 34.9 (May 1992), pp. 1005–1010.

[91] Marieke Krol, Werner Brouwer, and Frans Rutten. "Productivity Costs in Economic Evaluations: Past, Present, Future". In: *PharmacoEconomics* 31.7 (Apr. 2013), pp. 537–549.

[92] Amy L. Neftzger and Shannon Walker. "Measuring Productivity Loss Due to Health: A Multi-Method Approach". In: *Journal of Occupational and Environmental Medicine* 52.5 (May 2010), pp. 486–494.

[93] Macmillan Cancer Support. "Cancer's Hidden Price Tag: revealing the costs behind the illness". 2012.

[94] S. Byford and J. Raftery. "Economics notes: Perspectives in economic evaluation". In: *BMJ* 316.7143 (May 1998), pp. 1529–1530.

[95] J. McRae et al. "Cost Of Illness Studies From the Patient's Perspective: Study Design, Characteristics, and Costs". In: *Value in Health* 19.3 (May 2016).

[96] William E. Barlow. "Overview of Methods to Estimate the Medical Costs of Cancer". In: *Medical Care* 47.Supplement (July 2009), S33–S36.

[97] Ruth Etzioni, Nicole Urban, and Mary Baker. "Estimating the costs attributable to a disease with application to ovarian cancer". In: *Journal of Clinical Epidemiology* 49.1 (Jan. 1996), pp. 95–103.

[98] Shane Frederick, George Loewenstein, and Ted O'donoghue. "Time discounting and time preference: A critical review". In: (2002).

[99] Salvador Cruz Rambaud, María J. Muñoz Torrecillas, and Taiki Takahashi. "Observed and Normative Discount Functions in Addiction and other Diseases". In: *Frontiers in Pharmacology* 0.JUN (2017), p. 416.

[100] Gary S. Becker and Kevin M. Murphy. "A Theory of Rational Addiction". In: *https://doi.org/10.1086/261558* 96.4 (Oct. 2015), pp. 675–700.

[101] Warren K. Bickel et al. "Excessive discounting of delayed reinforcers as a trans-disease process contributing to addiction and other disease-related vulnerabilities: Emerging evidence". In: *Pharmacology & Therapeutics* 134.3 (June 2012), pp. 287–297.

[102] Jan Abel Olsen. "On what basis should health be discounted?"
In: *Journal of Health Economics* 12.1 (Apr. 1993), pp. 39–53.

[103] Olivier Godard. "Time discounting and long-run issues: the controversy raised by the Stern Review of the economics of climate change".
In: *OPEC Energy Review* 33.1 (Mar. 2009), pp. 1–22.

[104] Stéphane Hallgatte. "A proposal for a new prescriptive discounting scheme: The intergenerational discount rate". In: (2008).

[105] Mauro Laudicella et al. "Cost of care for cancer patients in England: evidence from population-based patient-level data".
In: *British Journal of Cancer* 114.11 (May 2016), pp. 1286–1292.

[106] James D. Lubitz and Gerald F. Riley.
"Trends in Medicare Payments in the Last Year of Life".
In: *New England Journal of Medicine* 328.15 (Apr. 1993), pp. 1092–1096.

[107] ISD Scotland. *National Data Catalogue | National Datasets*. 2021.
URL: https://www.ndc.scot.nhs.uk/National-Datasets/data.asp?SubID=23 (visited on 11/07/2021).

[108] Claudia Geue et al. "Spoilt For Choice: Implications Of Using Alternative Methods Of Costing Hospital Episode Statistics".
In: *Health Economics* 21.10 (Sept. 2011), pp. 1201–1216.

[109] Joanna C. Thorn et al.
"Validating the use of Hospital Episode Statistics data and comparison of costing methodologies for economic evaluation: an end-of-life case study from the Cluster randomised triAl of PSA testing for Prostate cancer (CAP)".
In: *BMJ Open* 6.4 (Apr. 2016), e011063.

[110] Sue Llewellyn et al.
"Patient-level information and costing systems (PLICSs): a mixed-methods study of current practice and future potential for the NHS health economy".
In: *Health Services and Delivery Research* 4.31 (Oct. 2016), pp. 1–156.

[111] E. Sopina et al.
"Long-term medical costs of Alzheimer's disease: matched cohort analysis".
In: *European Journal of Health Economics* (2018).

[112] Claire de Oliveira et al. "Trends in use and cost of initial cancer treatment in Ontario: a population-based descriptive study".
In: *CMAJ Open* 1.4 (Dec. 2013), E151–E158.

[113] Claire de Oliveira et al.
"Phase-specific and lifetime costs of cancer care in Ontario, Canada".
In: *BMC Cancer* 16.1 (Oct. 2016), p. 809.

[114] David E. Card et al.
"Expanding Access to Administrative Data for Research in the United States".
In: *SSRN Electronic Journal* (2010).

[115] Matthew Woollard. "Administrative Data: Problems and Benefits. A perspective from the United Kingdom". In: *Facing the Future: European Research Infrastructures for the Humanities and Social Sciences*. 2014.

[116] L. Einav and J. Levin. "Economics in the age of big data".
In: *Science* 346.6210 (Nov. 2014), p. 1243089.

[117] C.L. Philip Chen and Chun-Yang Zhang. "Data-intensive applications, challenges, techniques and technologies: A survey on Big Data".
In: *Information Sciences* 275 (Aug. 2014), pp. 314–347.

[118] Jane M. Geraci et al. "Comorbid disease and cancer: The need for more relevant conceptual models in health services research".
In: *Journal of Clinical Oncology* 23.30 (2005), pp. 7399–7404.

[119] Björn Stollenwerk et al. "Cost-of-illness studies based on massive data: a prevalence-based, top-down regression approach."
In: *The European journal of health economics : HEPAC : health economics in prevention and care* 17.3 (Apr. 2016), pp. 235–244.

[120] Eric I. Benchimol et al. "The Reporting of studies Conducted using Observational Routinely-collected health Data (RECORD) Statement".
In: *PLOS Medicine* 12.10 (Oct. 2015), e1001885.

[121] E.H. Morrato, M. Elias, and C.A. Gericke.
"Using population-based routine data for evidence-based health policy decisions: lessons from three examples of setting and evaluating national health policy in Australia, the UK and the USA".
In: *Journal of Public Health* 29.4 (Dec. 2007), pp. 463–471.

[122] Luc Rocher, Julien M. Hendrickx, and Yves Alexandre de Montjoye.
"Estimating the success of re-identifications in incomplete datasets using generative models".
In: *Nature Communications 2019 10:1* 10.1 (July 2019), pp. 1–9.

[123] Nick Triggle. *Care.data: How did it go so wrong? - BBC News*. 2014. URL: https://www.bbc.co.uk/news/health-26259101 (visited on 06/03/2021).

[124] Madhumita Murgia. *NHS hit by legal threat over GP data 'grab'*. 2021. URL: https://www.ft.com/content/a13225c8-b618-4ee4-ae16-a0e26829bc7b (visited on 06/04/2021).

[125] Editorial Board. *Plans to share NHS data must be reconsidered*. 2021. URL: https://www.ft.com/content/e5fbaf09-34f5-4a08-8d3a-7fbc6e9e3c44 (visited on 06/04/2021).

[126] UK Health Data Research Alliance. *Trusted Research Environments (TRE)*. Tech. rep. URL: https://www.theguardian.com/technology/2020/feb/08/fears-over-sale-anonymous-nhs-patient-data.

[127] Mackenzie Graham et al. "Trust and the Goldacre Review: why trusted research environments are not about trust". In: *Journal of Medical Ethics* 0 (Aug. 2022), pp. 1–4.

[128] Paul Affleck et al. "Trusted research environments are definitely about trust". In: *Journal of Medical Ethics* (2022).

[129] OpenSAFELY. *OpenSAFELY: Home*. 2021. URL: https://www.opensafely.org/ (visited on 06/14/2021).

[130] University of Edinburgh. *DataLoch*. 2021. URL: https://www.ed.ac.uk/usher/dataloch (visited on 05/18/2021).

[131] European Commission. *European Open Science Cloud (EOSC)*. 2021. URL: https://ec.europa.eu/info/research-and-innovation/strategy/goals-research-and-innovation-policy/open-science/european-open-science-cloud-eosc%7B%5C_%7Den (visited on 05/18/2021).

[132] Australian Research Data Commons. *ARDC – Australian Research Data Commons*. 2021. URL: https://ardc.edu.au/ (visited on 05/18/2021).

[133] Borislava Mihaylova et al. "Review of statistical methods for analysing healthcare resources and costs". In: *Health Economics* 20.8 (Aug. 2010), pp. 897–916.

[134] Peter C. Austin, William A. Ghali, and Jack V. Tu. "A comparison of several regression models for analysing cost of CABG surgery". In: *Statistics in Medicine* 22.17 (2003), pp. 2799–2815.

[135] Matthew P. Banegas et al. "Medical Care Costs Associated With Cancer in Integrated Delivery Systems". In: *Journal of the National Comprehensive Cancer Network* 16.4 (Apr. 2018), pp. 402–410.

[136] Martin J. Buxton et al. "Modelling in Ecomomic Evaluation: An Unavoidable Fact of Life". In: *Health Economics* (1997), pp. 217–227.

[137] W.M. Hart, C. Espinos, and J. Rovira.
"A simulation model of the cost of the incidence of IDDM in Spain".
In: *Diabetologia* (1997), pp. 311–318.

[138] Frank A. Sonnenberg and J. Robert Beck.
"Markov Models in Medical Decision Making".
In: *Medical Decision Making* (1993).

[139] Malcolm Faddy, Nicholas Graves, and Anthony Pettitt.
"Modeling Length of Stay in Hospital and Other Right Skewed Data: Comparison of Phase-Type, Gamma and Log-Normal Distributions".
In: *Value in Health* 12.2 (Mar. 2009), pp. 309–314.

[140] Andrew M. Jones. *Models for Health Care*. Oxford University Press, July 2011.

[141] Jonathan Karnon. "Alternative decision modelling techniques for the evaluation of health care technologies: Markov processes versus discrete event simulation".
In: *Health Economics* 12.10 (2003), pp. 837–848.

[142] Gery P. Guy et al.
"Melanoma treatment costs: A systematic review of the literature, 1990-2011".
In: *American Journal of Preventive Medicine* 43.5 (Nov. 2012), pp. 537–545.

[143] Johan J. Polder et al.
"A Cross-National Perspective on Cost of Illness: A Comparison of Studies from the Netherlands, Australia, Canada, Germany, United Kingdom, and Sweden."
In: *European Journal of Health Economics* 6.3 (2005), pp. 223–232.

[144] Avi Cherla et al. "Cost-effectiveness of cancer drugs: Comparative analysis of the United States and England". In: *EClinicalMedicine* 29-30 (Dec. 2020).

[145] J. Smith-Palmer, C. Takizawa, and W. Valentine.
"Literature review of the burden of prostate cancer in Germany, France, the United Kingdom and Canada". In: *BMC Urology* 19.1 (Mar. 2019), p. 19.

[146] Joachim Marti et al. "The economic burden of cancer in the UK: a study of survivors treated with curative intent".
In: *Psycho-Oncology* 25.1 (June 2015), pp. 77–83.

[147] Charmaine S. Ng et al.
"Cost-of-illness studies of diabetes mellitus: A systematic review".
In: *Diabetes Research and Clinical Practice* 105.2 (Aug. 2014), pp. 151–163.

[148] Sun Hee Rim et al. "The impact of chronic conditions on the economic burden of cancer survivorship: a systematic review". In: *Expert Review of Pharmacoeconomics & Outcomes Research* 16.5 (Sept. 2016), pp. 579–589.

[149] Organisation for Economic Co-operation and Development.
Conversion rates - Purchasing power parities (PPP).

URL: https://data.oecd.org/conversion/purchasing-power-parities-ppp.htm (visited on 12/21/2022).

[150] Bank of England. *Inflation calculator*.
URL: https://www.bankofengland.co.uk/monetary-policy/inflation/inflation-calculator (visited on 12/21/2022).

[151] Tony Blakely et al.
"Health system costs for individual and comorbid noncommunicable diseases: An analysis of publicly funded health events from New Zealand".
In: *PLOS Medicine* 16.1 (Jan. 2019). Ed. by Aziz Sheikh, e1002716.

[152] Christopher J. L. Murray et al.
"UK health performance: findings of the Global Burden of Disease Study 2010."
In: *Lancet (London, England)*. Comment in: Lancet. 2013 Mar 23;381(9871):970-2 PMID: 23668562 [https://www.ncbi.nlm.nih.gov/pubmed/23668562] 381.9871 (2013), pp. 997–1020.

[153] Sacha Hilhorst and Alan Lockey.
Cancer Costs: A 'ripple effect' analysis of cancer's wider impact. Tech. rep. 2020. URL: www.demos.co.uk.

[154] B. Hanratty et al. "Making the most of routine data in palliative care research–a case study analysis of linked hospital and mortality data on cancer and heart failure patients in Scotland and Oxford."
In: *Palliative medicine* 22.6 (2008), pp. 744–749.

[155] Lisa Iversen, Shona Fielding, and Philip C Hannaford.
"Smoking in young women in Scotland and future burden of hospital admission and death: a nested cohort study." In: *The British journal of general practice : the journal of the Royal College of General Practitioners*. Erratum in: Br J Gen Pract. 2013 Sep;63(614):463 63.613 (2013), e523–33.

[156] J Karnon et al. "Health care costs for the treatment of breast cancer recurrent events: estimates from a UK-based patient-level analysis."
In: *British journal of cancer* 97.4 (2007), pp. 479–485.

[157] K. M. de Ligt et al. "Patient-reported health problems and healthcare use after treatment for early-stage breast cancer."
In: *Breast (Edinburgh, Scotland)* 46 (2019), pp. 4–11.

[158] Chittaranjan Andrade. "The Limitations of Online Surveys".
In: *Indian Journal of Psychological Medicine* 42.6 (Oct. 2020), pp. 575–576.

[159] ISD Scotland. *Scottish Cancer Registry*. 2020. URL: https://www.isdscotland.org/Health-Topics/Cancer/Scottish-Cancer-Registry.asp.

[160] ISD Scotland. *National Data Catalogue | National Datasets*. 2020. URL: https://www.ndc.scot.nhs.uk/National-Datasets/data.asp?SubID=9 (visited on 07/16/2020).

[161] ISD Scotland. *Assessment of SMR01 Data Scotland 2014-2015*. Tech. rep. 2014.

[162] D. H. Brewster et al. "Reliability of cancer registration data in Scotland, 1997". In: *European Journal of Cancer* 38.3 (Jan. 2002), pp. 414–417.

[163] D. Brewster, J. Crichton, and C. Muir. "How accurate are Scottish cancer registration data?" In: *British Journal of Cancer 1994 70:5* 70.5 (1994), pp. 954–959.

[164] David Brewster, Calum Muir, and Judith Crichton. "Registration of lung cancer in Scotland: an assessment of data accuracy based on review of medical records". In: *Cancer Causes and Control* 6.4 (July 1995), pp. 303–310.

[165] D. Brewster, C. Muir, and J. Crichton. "Registration of colorectal cancer in Scotland: An assessment of data accuracy based on review of medical records". In: *Public Health* 109.4 (July 1995), pp. 285–292.

[166] Gov.scot. *Scottish Index of Multiple Deprivation 2020*. URL: https://www.gov.scot/collections/scottish-index-of-multiple-deprivation-2020 (visited on 11/08/2022).

[167] Harindra C. Wijeysundera et al. "Techniques for estimating health care costs with censored data: An overview for the health services researcher". In: *ClinicoEconomics and Outcomes Research* 4.1 (2012), pp. 145–155.

[168] Judith D. Willett and John B. Singer. *Doing data analysis with proportional hazards models: model building, interpretation and diagnosis*. Tech. rep. Apr. 1988.

[169] D. K. Blough and S. D. Ramsey. "Using Generalized Linear Models to Assess Medical Care Costs". In: *Health Services and Outcomes Research Methodology 2000 1:2* 1.2 (2000), pp. 185–202.

[170] Marcelo Coca Perraillon. *Cost data and Generalized Linear Models*. URL: https://clas.ucdenver.edu/marcelo-perraillon/sites/default/files/attached-files/week%7B%5C_%7D7%7B%5C_%7Dglm%7B%5C_%7Dand%7B%5C_%7Dcosts%7B%5C_%7Dperraillon.pdf (visited on 02/02/2022).

[171] Stata.com. *Stcox PH-assumption tests-Tests of proportional-hazards assumption*. 2021. URL:

https://www.stata.com/manuals/ststcoxph-assumptiontests.pdf (visited on 03/31/2021).

[172] Pamela Farley Short, John R. Moran, and Rajeshwari Punekar. "Medical expenditures of adult cancer survivors aged <65 years in the United States". In: *Cancer* (2011).

[173] Laurent Molinier et al. "Cost study of the clinical management of prostate cancer in France: Results on the basis of population-based data". In: *European Journal of Health Economics* 12.4 (Aug. 2011), pp. 363–371.

[174] David Steel and Jonathan Cylus. *United Kingdom (Scotland) Health system review Health Systems in Transition*. Tech. rep.

[175] Jaya S. Khushalani et al. "Systematic review of healthcare costs related to mental health conditions among cancer survivors". In: *Expert Review of Pharmacoeconomics and Outcomes Research* (2018).

[176] Lou-Ching Kuo et al. "End-of-Life Health Care Utilization Between Chronic Obstructive Pulmonary Disease and Lung Cancer Patients." In: *Journal of pain and symptom management* 57.5 (2019), pp. 933–943.

[177] Lucie Kutikova et al. "The economic burden of lung cancer and the associated costs of treatment failure in the United States." In: *Lung cancer (Amsterdam, Netherlands)* 50.2 (2005), pp. 143–154.

[178] Maria Pisu et al. "Costs of cancer along the care continuum: What we can expect based on recent literature." In: *Cancer* 124.21 (2018), pp. 4181–4191.

[179] Tony Blakely et al. "Patterns of cancer care costs in a country with detailed individual data." In: *Medical care*. Erratum in: Med Care. 2015 Jun;53(6):560 53.4 (2015), pp. 302–309.

[180] Kathleen Lang et al. "Survival, healthcare resource use and costs among stage IV ER + breast cancer patients not receiving HER2 targeted therapy: a retrospective analysis of linked SEER-Medicare data." In: *BMC health services research* 14 (2014), p. 298.

[181] Amresh D. Hanchate et al. "Longitudinal patterns in survival, comorbidity, healthcare utilization and quality of care among older women following breast cancer diagnosis." In: *Journal of general internal medicine* 25.10 (2010), pp. 1045–1050.

[182] Rahul Khanna et al. "Prevalence, healthcare utilization, and costs of breast cancer in a state Medicaid fee-for-service program." In: *Journal of women's health (2002)* 20.5 (2011), pp. 739–747.

[183] Kathleen Lang et al. "Trends in healthcare utilization among older Americans with colorectal cancer: a retrospective database analysis."
In: *BMC health services research* 9 (2009), p. 227.

[184] Xue Song et al. "Cost of illness in patients with metastatic colorectal cancer."
In: *Journal of medical economics* 14.1 (2011), pp. 1–9.

[185] Ravishankar Jayadevappa et al.
"Medical care cost of patients with prostate cancer."
In: *Urologic oncology* 23.3 (2005), pp. 155–162.

[186] Stella Chang et al.
"Estimating the cost of cancer: results on the basis of claims data analyses for cancer patients diagnosed with seven types of cancer during 1999 to 2000."
In: *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 22.17 (2004), pp. 3524–3530.

[187] David G. Kleinbaum and Mitchel Klein.
"Analysis of Matched Data Using Logistic Regression". In:
Statistics for Biology and Health. New York: Springer New York, 2010,
pp. 389–428.

[188] Karen Barnett et al. "Epidemiology of multimorbidity and implications for health care, research, and medical education: A cross-sectional study".
In: *The Lancet* 380.9836 (July 2012), pp. 37–43.

[189] Institute for Health Metrics and Evaluation.
The Lancet: Latest global disease estimates reveal perfect storm of rising chronic diseases and public health failures fuelling COVID-19 pandemic. 2020.
URL: https://www.healthdata.org/news-release/lancet-latest-global-disease-estimates-reveal-perfect-storm-rising-chronic-diseases-and (visited on 08/15/2022).

[190] Theo Vos et al.
"Global burden of 369 diseases and injuries in 204 countries and territories, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019".
In: *The Lancet* 396 (2020), pp. 1204–1222.

[191] Harmon Eyre et al. "Preventing cancer, cardiovascular disease, and diabetes: a common agenda for the American Cancer Society, the American Diabetes Association, and the American Heart Association."
In: *Circulation* 109.25 (2004), pp. 3244–3255.

[192] World Health Organisation. *Cardiovascular diseases (CVDs)*. 2021.
URL: https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds) (visited on 04/22/2021).

[193] Kevin D. Shield, Charles Parry, and Jürgen Rehm.
"Chronic diseases and conditions related to alcohol use".
In: *Alcohol Research: Current Reviews* 35.2 (2013), pp. 155–171.

[194] C. Kreatsoulas, S. S. Anand, and S. V. Subramanian.
"An emerging double burden of disease: the prevalence of individuals with cardiovascular disease and cancer".
In: *Journal of Internal Medicine* 275.5 (May 2014), pp. 494–505.

[195] Y. Claire Wang et al. *Health and economic burden of the projected obesity trends in the USA and the UK*. Aug. 2011.

[196] John B. Dixon. "The effect of obesity on health outcomes".
In: *Molecular and Cellular Endocrinology* 316.2 (Mar. 2010), pp. 104–108.

[197] Corinne R. Leach et al. "The complex health profile of long-term cancer survivors: prevalence and predictors of comorbid conditions".
In: *Journal of Cancer Survivorship* (2015).

[198] Charlie Foster and Steven Allender.
Costing the burden of ill health related to physical inactivity for Scotland.
Tech. rep. 2012.
URL: https://www.healthscotland.com/documents/6262.aspx.

[199] Joceline Pomerleau, Karen Lock, and Martin McKee. "The burden of cardiovascular disease and cancer attributable to low fruit and vegetable intake in the European Union: differences between old and new Member States".
In: *Public Health Nutrition* 9.5 (Aug. 2006), pp. 575–583.

[200] Annemarie A. Uijen and Eloy H. van de Lisdonk.
"Multimorbidity in primary care: Prevalence and trend over the last 20 years".
In: *European Journal of General Practice* 14.sup1 (Jan. 2008), pp. 28–32.

[201] Cristiana Abbafati et al.
"Global burden of 87 risk factors in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019".
In: *The Lancet* 396.10258 (Oct. 2020), pp. 1223–1249.

[202] Manuel B. Huber et al. "Excess Costs of Comorbidities in Chronic Obstructive Pulmonary Disease: A Systematic Review".
In: *PLOS ONE* 10.4 (Apr. 2015). Ed. by Henrik Watz, e0123292.

[203] Gery P. Jr. Guy et al. "Economic Burden of Chronic Conditions Among Survivors of Cancer in the United States."
In: *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 35.18 (2017), pp. 2053–2061.

[204] Hude Quan et al. "Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data".
In: *Medical Care* 43.11 (Nov. 2005), pp. 1130–1139.

[205] ISD Scotland. *ISD Services | Data Support and Monitoring | SMR Completeness | ISD Scotland*. 2022.
URL: https://www.isdscotland.org/products-and-Services/Data-Support-and-Monitoring/SMR-Completeness/ (visited on 08/19/2022).

[206] Colin Cameron and Douglas Miller.
"A Practitioner's Guide to Cluster-Robust Inference".
In: *Journal of Human Resources* 50.2 (2015), pp. 317–372.

[207] Federico Belotti et al. "twopm: Two-part models".
In: *The Stata Journal* 15.1 (2015), pp. 3–20.

[208] Joseph L. Dieleman et al. "Adjusting health spending for the presence of comorbidities: an application to United States national inpatient data."
In: *Health economics review* 7.1 (Aug. 2017), p. 30.

[209] Jin Choi et al. "Association of diabetes with frequency and cost of hospital admissions: a retrospective cohort study". In: *Canadian Medical Association Open Access Journal* 9.2 (Apr. 2021), E406–E412.

[210] Susan Morton and Martin Mckee.
Understanding the Health of Scotland's Population in an International Context.
Tech. rep. London School of Hygiene & Tropical Medicine, 2003.
URL: https://www.researchgate.net/publication/242666287.

[211] M. C. Liu et al. "Sensitive and specific multi-cancer detection and localization using methylation signatures in cell-free DNA".
In: *Annals of Oncology* 31.6 (June 2020), pp. 745–759.

[212] J. Maddams, M. Utley, and H. Møller.
"Projections of cancer prevalence in the United Kingdom, 2010-2040".
In: *British Journal of Cancer* (2012).

[213] Angela G. E. M. de Boer et al. "Cancer Survivors and Unemployment".
In: *JAMA* 301.7 (Feb. 2009), p. 753.

[214] Department of Work and Pensions.
Proposed new timetable for State Pension age increases. 2017.
URL: https://www.gov.uk/government/news/proposed-new-timetable-for-state-pension-age-increases (visited on 06/26/2019).

[215] J. Weis, U. Koch, and M. Geldsetzer. "Changes in occupational status following cancer. An empirical study on occupational rehabilitation."
In: *Sozial- und Praventivmedizin* (1992).

[216] Anna Zajacova et al. "Employment and income losses among cancer survivors: Estimates from a national longitudinal survey of American families". In: *Cancer* 121.24 (Dec. 2015), pp. 4425–4432.

[217] Cathy J. Bradley et al. "Short-term effects of breast cancer on labor market attachment: Results from a longitudinal study". In: *Journal of Health Economics* (2005).

[218] Cathy J. Bradley et al. "Employment and cancer: Findings from a longitudinal study of breast and prostate cancer survivors". In: *Cancer Investigation* (2007).

[219] David Candon. "The effects of cancer on older workers in the English labour market". In: *Economics & Human Biology* 18 (July 2015), pp. 74–84.

[220] Maria D. Thomson and Laura A. Siminoff. "Managing work and cancer treatment: Experiences among survivors of hematological cancer". In: *Cancer* 124.13 (July 2018), pp. 2824–2831.

[221] Pamela Farley Short, Joseph J. Vasey, and Kaan Tunceli. "Employment pathways in a large cohort of adult cancer survivors". In: *Cancer* (2005).

[222] John R. Moran, Pamela Farley Short, and Christopher S. Hollenbeak. "Long-term employment effects of surviving cancer". In: *Journal of Health Economics* 30.3 (May 2011), pp. 505–514.

[223] Pamela Farley Short, Joseph J. Vasey, and Rhonda BeLue. "Work disability associated with cancer survivorship and other chronic conditions". In: *Psycho-Oncology* (2008).

[224] Elizabeth Maunsell et al. "Work situation after breast cancer: Results from a population-based study". In: *Journal of the National Cancer Institute* (2004).

[225] Sung-Hee Jeon. "The Long-Term Effects of Cancer on Employment and Earnings". In: *Health Economics* 26.5 (May 2017), pp. 671–684.

[226] Steffen Torp et al. "Change in employment status of 5-year cancer survivors". In: *European Journal of Public Health* (2013).

[227] Taina Taskila-Åbrandt et al. "The impact of education and occupation on the employment status of cancer survivors". In: *European Journal of Cancer* (2004).

[228] Vicki S. Helgeson and Patricia L. Tomich. "Surviving cancer: A comparison of 5-year disease-free breast cancer survivors with healthy women". In: *Psycho-Oncology* 14.4 (Apr. 2005), pp. 307–317.

[229] Cathy J. Bradley et al.
"Employment outcomes of men treated for prostate cancer".
In: *Journal of the National Cancer Institute* (2005).

[230] Phyllis Butow et al. "Return to work after a cancer diagnosis: a meta-review of reviews and a meta-synthesis of recent qualitative studies".
In: *Journal of Cancer Survivorship* (Dec. 2019).

[231] Evelien R. Spelten, Mirjam A G Sprangers, and Jos H A M Verbeek.
"Factors reported to influence the return to work of cancer survivors: A literature review". In: *Psycho-Oncology* (2002).

[232] T. Taskila and M. L. Lindbohm.
"Factors affecting cancer survivors' employment and work ability".
In: *Acta Oncologica* (2007).

[233] Eskil Heinesen, Susumu Imai, and Shiko Maruyama.
"Employment, job skills and occupational mobility of cancer survivors".
In: *Journal of Health Economics* 58 (Mar. 2018), pp. 151–175.

[234] Eskil Heinesen et al.
"Return to work after cancer and pre-cancer job dissatisfaction".
In: *Applied Economics* 49.49 (Oct. 2017), pp. 4982–4998.

[235] Sarah E. Lewis, Maryam Doroudi, and K. Robin Yabroff. "Financial Hardship".
In: *Handbook of Cancer Survivorship* (2018), pp. 111–125.

[236] Diane Von Ah et al. "Work". In: *Handbook of Cancer Survivorship*.
Cham: Springer International Publishing, 2018, pp. 227–242.

[237] John F. Steiner et al. "The impact of physical and psychosocial factors on work characteristics after cancer". In: *Psycho-Oncology* (2008).

[238] Karen M. Mustian et al. "Fatigue".
In: *Handbook of Cancer Survivorship* (2018), pp. 129–144.

[239] Julienne E. Bower.
"Cancer-related fatigue–mechanisms, risk factors, and treatments."
In: *Nature reviews. Clinical oncology* (2014).

[240] E. R. Spelten et al. "Cancer, fatigue and the return of patients to work - A prospective cohort study". In: *European Journal of Cancer* (2003).

[241] Michael Nekhlyudov, Larissa and Feuerstein. *Handbook of Cancer Survivorship*. Ed. by Michael Nekhlyudov, Larissa and Feuerstein. 2nd. 2018.

[242] M. K. Lee et al. "Employment status and work-related difficulties in stomach cancer survivors compared with the general population".
In: *British Journal of Cancer* (2008).

[243] Marie-Hélène Savard and Josée Savard. "Sleep".
In: *Handbook of Cancer Survivorship*.
Cham: Springer International Publishing, 2018, pp. 243–264.

[244] M.F. Roizen. "The Economic Burden of Insomnia: Direct and Indirect Costs for Individuals with Insomnia Syndrome, Insomnia Symptoms, and Good Sleepers".
In: *Yearbook of Anesthesiology and Pain Management* (2010).

[245] Peter Herschbach et al.
"Distress in cancer patients: who are the main groups at risk?"
In: *Psycho-Oncology* (Dec. 2019), pon.5321.

[246] A. Carrato et al. "A Systematic Review of the Burden of Pancreatic Cancer in Europe: Real-World Impact on Survival, Quality of Life and Costs".
In: *Journal of Gastrointestinal Cancer* 46.3 (Sept. 2015), pp. 201–211.

[247] Linda E. Carlson, Kirsti Toivonen, and Peter Trask. "Distress".
In: *Handbook of Cancer Survivorship*.
Cham: Springer International Publishing, 2018, pp. 145–166.

[248] Victor T. Chang and Neena Kapoor-Hintzen. "Pain".
In: *Handbook of Cancer Survivorship*.
Cham: Springer International Publishing, 2018, pp. 167–195.

[249] Jonathan Sussman, Eva Grunfeld, and Craig C. Earle. "Quality Care".
In: *Handbook of Cancer Survivorship*. 2nd.
Cham: Springer International Publishing, 2018, pp. 49–69.

[250] Robert J. Ferguson et al. "Cognitive Dysfunction".
In: *Handbook of Cancer Survivorship*.
Cham: Springer International Publishing, 2018, pp. 199–225.

[251] Tara O. Henderson, Kirsten K. Ness, and Harvey Jay Cohen.
"Accelerated Aging among Cancer Survivors: From Pediatrics to Geriatrics".
In: *American Society of Clinical Oncology Educational Book* (2014).

[252] A. G.E.M. De Boer et al. "Work ability and return-to-work in cancer patients".
In: *British Journal of Cancer* (2008).

[253] Leida M. Lamers et al.
"The relationship between productivity and health-related quality of life: An empirical exploration in persons with low back pain".
In: *Quality of Life Research* (2005).

[254] Wiji Arulampalam, Paul Gregg, and Mary Gregory. "Unemployment Scarring".
In: *The Economic Journal* (2001).

[255] Helen Blumen, Kathryn Fitch, and Vincent Polkus. "Comparison of Treatment Costs for Breast Cancer, by Tumor Stage and Type of Service". In: *American Health & Drug Benefits* 9.1 (Feb. 2016), p. 23.

[256] National Institute for Health and Care Excellence. *NICE health technology evaluations: the manual*. Tech. rep. 2022. URL: www.nice.org.uk/process/pmg36.

[257] Tanya P. Garcia and Karen Marder. "Statistical Approaches to Longitudinal Data Analysis in Neurodegenerative Diseases: Huntington's Disease as a Model". In: *Current Neurology and Neuroscience Reports* 17.2 (Feb. 2017), p. 14.

[258] D. B. Rubin. *Causal Inference*. Elsevier, Jan. 2010, pp. 66–71.

[259] Stephanie McFall and Nick Buck. "Understanding Society – The UK Household Longitudinal Survey: A Resource for Demographers". In: *Applied Demography and Public Health*. 2013, pp. 357–369.

[260] NatCen Social Research Institute for Social and Economic Affairs. "Understanding Society User Guide". In: *University of Essex* 1 (2019).

[261] Khalid M. Kamal et al. "A Systematic Review of the Effect of Cancer Treatment on Work Productivity of Patients and Caregivers". In: *Journal of Managed Care & Specialty Pharmacy* 23.2 (Feb. 2017), pp. 136–162.

[262] Ulf Seifart and Jan Schmielau. "Return to Work of Cancer Survivors". In: *Oncology Research and Treatment* 40.12 (2017), pp. 760–763.

[263] Esther Curnock, Alastair H. Leyland, and Frank Popham. "The impact on health of employment and welfare transitions for those receiving out-of-work disability benefits in the UK". In: *Social Science & Medicine* 162 (Aug. 2016), pp. 1–10.

[264] Andrew M. Jones et al. *Applied health economics*. 2007.

[265] Peter Lynn and Magda Borkowska. "Some Indicators of Sample Representativeness and Attrition Bias for BHPS and Understanding Society". 2018.

[266] Z. Fewell, G. Davey Smith, and J. A. C. Sterne. "The Impact of Residual and Unmeasured Confounding in Epidemiologic Studies: A Simulation Study". In: *American Journal of Epidemiology* 166.6 (June 2007), pp. 646–655.

[267] Juan M. Villa. "diff: Simplifying the estimation of difference-in-differences treatment effects". In: *Stata Journal* (2016).

[268] Jamie R. Daw and Laura A. Hatfield.
"Matching and Regression to the Mean in Difference-in-Differences Analysis".
In: *Health Services Research* 53.6 (Dec. 2018), pp. 4138–4156.

[269] Pilar García-Gómez, Andrew M. Jones, and Nigel Rice.
"Health effects on labour market exits and entries".
In: *Labour Economics* 17.1 (Jan. 2010), pp. 62–76.

[270] Dawn S. Stone et al.
"Young adult cancer survivors and work: a systematic review".
In: *Journal of Cancer Survivorship* 11.6 (Dec. 2017), pp. 765–781.

[271] Mark P. Mankiw, Gregory N.; Taylor. *Economics*. 4th. Cengage Learning, 2017.

[272] Department for Work and Pensions. *State Pension age Review 2023*.
URL: https://www.gov.uk/government/publications/state-pension-age-review-2023-government-report/state-pension-age-review-2023
(visited on 05/23/2023).

[273] M. Bartley. "Employment status, employment conditions, and limiting illness: prospective evidence from the British household panel survey 1991-2001". In: *Journal of Epidemiology & Community Health* 58.6 (June 2004), pp. 501–506.

[274] M. Bartley. "Unemployment and ill health: understanding the relationship." In: *Journal of Epidemiology & Community Health* 48.4 (Aug. 1994), pp. 333–337.

[275] Alan D. Lopez et al. "Global and regional burden of disease and risk factors, 2001: systematic analysis of population health data".
In: *The Lancet* 367.9524 (May 2006), pp. 1747–1757.

[276] Marlou E.C.L. van Broekhoven et al. "Illness perceptions and changes in lifestyle following a gynecological cancer diagnosis: A longitudinal analysis".
In: *Gynecologic Oncology* 145.2 (May 2017), pp. 310–318.

[277] Shih-Feng Cho et al. "Investigation of treatment pattern, medical resource utilization and demographic prognostic factors in older patients with non-Hodgkin lymphoma: A nationwide population-based study."
In: *Journal of geriatric oncology* 9.4 (2018), pp. 315–320.

[278] Lucie Kutikova et al. "Medical costs associated with non-Hodgkin's lymphoma in the United States during the first two years of treatment."
In: *Leukemia & lymphoma* 47.8 (2006), pp. 1535–1544.

[279] Michel van Agthoven et al.
"Cost determinants in aggressive non-Hodgkin's lymphoma."
In: *Haematologica* 90.5 (2005), pp. 661–671.

[280] David Stuckler et al. "Austerity and health: the impact in the UK and Europe".
In: *European Journal of Public Health* 27.suppl_4 (Oct. 2017), pp. 18–21.

[281] Karl Friston. "Ten ironic rules for non-statistical reviewers". In: *NeuroImage* 61.4 (July 2012), pp. 1300–1310.

[282] Michael Ingre. "Why small low-powered studies are worse than large high-powered studies and how to protect against "trivial" findings in research: Comment on Friston (2012)". In: *NeuroImage* 81 (Nov. 2013), pp. 496–498.

[283] John P. A. Ioannidis. "Why Most Discovered True Associations Are Inflated". In: *Epidemiology* 19.5 (Sept. 2008), pp. 640–648.

[284] Valérie Paris. "Health Systems Institutional Characteristics: A Survey of 29 OECD Countries". 2010.

[285] Camille Maringe et al. "The impact of the COVID-19 pandemic on cancer deaths due to delays in diagnosis in England, UK: a national, population-based, modelling study". In: *The Lancet Oncology* 21.8 (Aug. 2020), pp. 1023–1034.

[286] Halsey and Annie. *NHS backlogs and waiting times in England*. Tech. rep. National audit Office. URL: https://www.nao.org.uk/reports/managing-nhs-backlogs-and-waiting-times-in-england/.

[287] Norman E. Sharpless. "COVID-19 and cancer". In: *Science* 368.6497 (June 2020), p. 1290.

[288] Office for National Statistics. *LFS: Economic inactivity rate: UK: All: Aged 50-64*. URL: https://www.ons.gov.uk/employmentandlabourmarket/peoplenotinwork/economicinactivity/timeseries/lf2w/lms (visited on 11/02/2022).

[289] Scottish Government. *Cancer Action Plan 2023 to 2026*. Tech. rep. 2023. URL: http://www.gov.scot/publications/cancer-action-plan-scotland-2023-2026/.

[290] Public Health Scotland. *Evaluating the impact of minimum unit pricing for alcohol in Scotland: Final report*. Tech. rep. 2023. URL: www.publichealthscotland.scot.

Appendices

A Ovid Search Strategy for Cancer Costs

1 Resource Allocation/ec 618
2 Hospital Costs/sn 3474
3 "Cost of Illness"/ 30299
4 Health Services/ec, sn 11399
5 Long-Term Care/ec, og, sn 5864
6 Inpatients/sn 5010
7 Outpatients/sn 2834
8 Hospital Charges/ 3240
9 "Costs and Cost Analysis"/ec, sn, td 2098
10 Cost-Benefit Analysis/ec, sn 2927
11 Health Care Costs/ 42849
12 ((resource? or health?care) adj2 ("use" or "usage" or util* or consum*)).ti. 6491
13 direct costs.mp. 5368
14 or/1-13 108954
15 exp Neoplasms/co, de, dt, ec, ep, mo, nu, re, rt, rh, sn, su, tu, th 1933526
16 exp Lung Neoplasms/co, de, dt, ec, ep, mo, nu, re, rt, rh, sn, su, tu, th 135048
17 Breast Neoplasms/co, de, dt, ec, ep, mo, nu, re, rt, rh, sn, su, tu, th 162084
18 Colorectal Neoplasms/co, de, dt, ec, ep, mo, nu, re, rt, rh, sn, su, tu, th 50873
19 Prostatic Neoplasms/co, de, dt, ec, ep, mo, nu, re, rt, rh, sn, su, tu, th 74901
20 Kidney Neoplasms/co, de, dt, ec, ep, mo, nu, re, rt, rh, sn, su, tu, th 39879
21 Esophageal Neoplasms/co, de, dt, ec, ep, mo, nu, re, rt, rh, sn, su, tu, th 35228
22 (head and neck cancer).mp. 24616
23 Lymphoma, Non-Hodgkin/co, de, dt, ec, ep, mo, nu, re, rt, rh, sn, su, tu, th 19365
24 Skin Neoplasms/co, de, dt, ec, ep, mo, nu, re, rt, rh, sn, su, tu, th 56779
25 Urinary Bladder Neoplasms/co, de, dt, ec, ep, mo, nu, re, rt, rh, sn, su, tu, th 32513
26 (cancer* or malignan* or neoplas* or melanoma* or carcin*).mp. 3653264
27 or/15-26 3872042
28 14 and 27 11415
29 limit 28 to (english language and humans and last 15 years) 8422
30 limit 29 to "all adult (19 plus years)" 5177

SUPPLEMENTARY TERMS

("long?term" or trajector* or life?time or dynamic? or "incidence cost?").mp. [mp=title, abstract, original title, subject heading word, floating sub-heading word, keyword heading word, protocol supplementary concept word, unique identifier, synonyms]

(Scotland? or Scottish).mp. [mp=title, abstract, original title, subject heading word, floating sub-heading word, keyword heading word, protocol supplementary concept word, unique identifier, synonyms]

B Tabular Trajectories of Healthcare Costs

Table B.1.: Total costs by year

	Mean (£) Year from diagnosis								
Cancer type	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**Total**
Trachea, bronchus & Lung	15,257	2,164	969	676	518	437	376	289	20,686
Breast	10,498	3,827	3,122	2,582	2,421	2,427	2,278	1,959	29,114
Colorectal	16,585	4,632	3,017	2,440	2,019	1,916	1,774	1,574	33,957
Prostate	6,651	3,961	3,473	2,894	2,720	2,546	2,375	2,272	26,892
Head & neck	22,505	4,582	3,523	2,836	2,326	2,230	2,068	1,964	42,034
Malignant melanoma of skin	5,148	2,754	2,116	1,949	1,874	1,970	1,603	1,803	19,217
Kidney	12,119	3,310	2,441	2,382	2,312	1,876	1,816	1,841	28,097
Non-Hodgkin lymphoma	24,074	5,281	3,847	3,560	3,027	3,063	2,543	2,277	47,672
Oesophagus	23,465	3,577	1,358	749	670	419	331	294	30,863
Bladder	17,188	5,128	2,962	2,263	1,761	1,680	1,544	1,467	33,993
All other cancers	16,594	3,940	2,589	1,958	1,665	1,496	1,313	1,163	30,718

Notes: All participants were represented in all years. Years are relative to the diagnosis.

Table B.2.: Total costs by phase-of-care

	Initial			Continuous			End-of-life		
Cancer type	**Mean (£)**	**95% CI (low/high)**		**Mean (£)**	**95% CI (low/high)**		**Mean (£)**	**95% CI (low/high)**	
Trachea, bronchus & Lung	15,775	15,308	16,241	15,213	14,212	16,213	12,128	11,873	12,384
Breast	10,077	9,882	10,271	16,058	15,573	16,543	15,011	14,406	15,615
Colorectal	15,839	15,517	16,162	18,622	17,994	19,251	15,409	14,919	15,899
Prostate	5,530	5,376	5,683	16,894	16,386	17,402	14,902	14,391	15,414
Head & neck	21,079	20,281	21,876	20,917	19,390	22,444	18,316	17,349	19,284
Malignant melanoma of skin	4,586	4,312	4,860	11,603	10,833	12,373	13,979	12,855	15,103
Kidney	10,275	9,710	10,839	18,395	16,905	19,885	13,889	13,040	14,738
Non-Hodgkin lymphoma	22,266	21,002	23,530	24,481	22,666	26,296	24,746	22,890	26,602
Oesophagus	22,569	21,126	24,012	13,966	12,086	15,846	18,569	17,704	19,435
Bladder	14,580	13,785	15,375	22,185	20,593	23,778	16,410	15,553	17,267
All other cancers	16,120	15,728	16,512	20,583	19,805	21,361	16,260	15,897	16,623

Notes: CI = confidence interval. Only participants who entered a phase were represented in that phase. All costs are undiscounted at 2018 price levels.

Table B.3.: Monthly total costs by phase-of-care

	Initial			Continuous			End-of-life		
Cancer type	**Mean (£)**	**95% CI (low/high)**		**Mean (£)**	**95% CI (low/high)**		**Mean (£)**	**95% CI (low/high)**	
Trachea, bronchus & Lung	2,173	2,095	2,252	441	405	478	3,298	3,233	3,362
Breast	883	865	901	280	269	292	1,763	1,694	1,832
Colorectal	1,568	1,530	1,606	405	382	427	2,617	2,540	2,694
Prostate	527	509	545	323	310	336	1,807	1,735	1,879
Head & neck	2,146	2,051	2,241	391	364	419	2,589	2,452	2,726
Malignant melanoma of skin	423	396	450	197	179	214	1,535	1,406	1,664
Kidney	1,087	984	1,189	346	314	378	2,763	2,596	2,930
Non-Hodgkin lymphoma	2,112	1,984	2,240	448	395	500	4,294	4,025	4,564
Oesophagus	2,883	2,709	3,056	508	397	619	3,254	3,125	3,384
Bladder	1,574	1,486	1,662	472	422	521	2,523	2,378	2,667
All other cancers	1,755	1,707	1,804	463	441	486	3,543	3,480	3,605

Notes: CI = confidence interval. Only participants who entered a phase were represented in that phase. Total costs were the sum of inpatient/daycase, outpatient and prescriptions. All costs are undiscounted at 2018 price levels.

C Ovid Search Strategy for the Association Between Cancer and In-work Productivity

1. productivity.mp. [mp=title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, organism supplementary concept word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]

2. (indirect cost$ or indirect loss$).mp. [mp=title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, organism supplementary concept word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]

3. Work/ or Work Capacity Evaluation/ or Employment/

4. mortality cost$.mp. [mp=title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, organism supplementary concept word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]

5. Work impairment.mp. [mp=title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, organism supplementary concept word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]

6. (Employment loss$ or employment outcome$).mp. [mp=title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, organism supplementary concept word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]

7. Absenteeism.mp. [mp=title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, organism supplementary concept word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]

8. Presenteeism.mp. [mp=title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, organism supplementary concept word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]

9. Workforce participation.mp. [mp=title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, organism supplementary concept word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]

10. Unemploy$.mp. [mp=title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, organism supplementary concept word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]

11. Work ability.mp. [mp=title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, organism supplementary concept word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]

12. Sick leave.mp. [mp=title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, organism supplementary concept word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]

13. Missed work.mp. [mp=title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, organism supplementary concept word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]

14. (Time off work or time away from work).mp. [mp=title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, organism supplementary concept word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]

15. (friction cost$ or FCM).mp. [mp=title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, organism supplementary concept word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]

16. (human capital method or HCM).mp. [mp=title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, organism supplementary concept word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]

17. 1 or 2 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16

18. Neoplasms/co, de, dt, ec, ep, gd, hi, mo, ph, pc, px, rt, sn, su, th [Complications, Drug Effects, Drug Therapy, Economics, Epidemiology, Growth & Development, History, Mortality, Physiology, Prevention & Control, Psychology, Radiotherapy, Statistics & Numerical Data, Surgery, Therapy]

19. cancer.mp. [mp=title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, organism supplementary concept word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]

20. neoplasm.mp. [mp=title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, organism supplementary concept word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]

21. malignant tumor.mp. [mp=title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, organism supplementary concept word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]

22. oncology.mp. [mp=title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, organism supplementary concept word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]

23. 18 or 19 or 20 or 21 or 22

24. 17 and 23

25. limit 24 to humans

26. limit 25 to "all adult (19 plus years)"

27. limit 26 to english language

28. limit 27 to yr="2000 - 2019"

29. limit 28 to ("economics (best balance of sensitivity and specificity)" or "qualitative (best balance of sensitivity and specificity)" or "costs (best balance of sensitivity and specificity)")

D Additional Results for Analysis A in Chapter 6

Table D.1.: Multivariable logistic regression on working with comorbidities by time for Analysis A

Explanatory Variable	Odds Ratio	Std. Error	p	Odds Ratio 95% CI	
Cancer 1-2 years	0.659	0.136	0.043	0.439	0.988
Cancer 3-5 years	0.652	0.121	0.021	0.454	0.937
Cancer 6-10 years	0.798	0.143	0.209	0.561	1.135
Cancer >10 years	0.891	0.159	0.515	0.628	1.263
Asthma	0.893	0.048	0.034	0.804	0.992
Arthritis	0.523	0.027	<0.001	0.472	0.579
Congestive heart failure	0.427	0.124	0.003	0.241	0.755
Coronary heart failure	0.689	0.110	0.020	0.503	0.942
Angina	0.516	0.069	<0.001	0.397	0.670
Heart attack or myocardial infarction	0.580	0.082	<0.001	0.439	0.766
Stroke	0.343	0.051	<0.001	0.257	0.460
Emphysema	0.404	0.100	<0.001	0.249	0.658
Hyperthyroidism or an over-active thyroid	0.925	0.157	0.645	0.664	1.289
Hypothyroidism or an under-active thyroid	0.937	0.089	0.495	0.778	1.129
Chronic bronchitis	0.577	0.064	<0.001	0.464	0.716
Any kind of liver condition	0.586	0.070	<0.001	0.463	0.741
Diabetes	0.491	0.033	<0.001	0.431	0.561
Epilepsy	0.356	0.045	<0.001	0.278	0.457
High blood pressure	0.866	0.042	0.003	0.787	0.953
Clinical depression	0.331	0.019	<0.001	0.297	0.369
None of these	0.931	0.048	0.172	0.841	1.031
age	1.343	0.029	<0.001	1.287	1.401
age^2	0.996	0.000	<0.001	0.996	0.997
30-44 (base)	1				
45-54 years old	1.551	0.099	<0.001	1.369	1.757
55-64 years old	1.730	0.202	<0.001	1.376	2.175
Male (base)	1				
Female	0.558	0.018	<0.001	0.524	0.595
No higher qualification (base)	1				
Higher level qualification	2.049	0.067	<0.001	1.922	2.185
Not married (base)	1				
Married or civil partnership	1.423	0.044	<0.001	1.339	1.512
UK white (base)	1				
Non UK white	0.424	0.015	<0.001	0.396	0.454
Parent working aged 14 (base)	1				
No parents working aged 14	0.533	0.027	<0.001	0.482	0.589
_constant	0.017	0.008	<0.001	0.007	0.041
			area under ROC curve =	0.7375	

Notes: SF-12 = Short-Form 12, ROC = receiver operating characteristic. CI = confidence interval.

Table D.2.: Weighted univariable and multivariable regression results for Analysis A

LOGISTIC MODELS		**UNIVARIABLE**		**MULTIVARIABLE**	
Outcome	**Variable**	**Odds Ratio**	***p***	**Odds Ratio**	***p***
Working	No cancer diagnosis	1	Reference	1	Reference
	Cancer at t=1-2 years	0.536	0.001	0.727	0.126
	Cancer at t=3-5 years	0.435	<0.001	0.528	<0.001
	Cancer at t=6-10 years	0.680	0.031	0.762	0.137
	Cancer at t>10 years	0.626	0.002	0.732	0.052
OLS MODELS		**UNIVARIABLE**		**MULTIVARIABLE**	
Outcome	**Variable**	**Coefficient**	***p***	**Coefficient**	***p***
Job hours	No cancer diagnosis	0	Reference	0	Reference
	Cancer at t=1-2 years	-0.317	0.798	0.404	0.714
	Cancer at t=3-5 years	-2.963	0.020	-1.592	0.149
	Cancer at t=6-10 years	-1.929	0.101	-0.778	0.457
	Cancer at t>10 years	-0.084	0.925	1.225	0.133
Monthly earnings (£)	No cancer diagnosis	1	Reference	0	Reference
	Cancer at t=1-2 years	182.503	0.390	315.111	0.076
	Cancer at t=3-5 years	-301.616	0.088	-198.819	0.190
	Cancer at t=6-10 years	-198.179	0.263	-201.314	0.183
	Cancer at t>10 years	-94.013	0.507	7.994	0.951
Log earnings	No cancer diagnosis	1	Reference	0	Reference
	Cancer at t=1-2 years	0.126	0.339	0.202	0.092
	Cancer at t=3-5 years	-0.251	0.084	-0.196	0.158
	Cancer at t=6-10 years	-0.093	0.397	-0.085	0.411
	Cancer at t>10 years	-0.015	0.864	0.043	0.596
Monthly income (£)	No cancer diagnosis	1	Reference	0	Reference
	Cancer at t=1-2 years	-135.296	0.201	-60.453	0.533
	Cancer at t=3-5 years	-337.067	<0.001	-235.128	0.002
	Cancer at t=6-10 years	11.166	0.944	80.732	0.571
	Cancer at t>10 years	-99.403	0.380	-11.083	0.917
Log income	No cancer diagnosis	1	Reference	0	Reference
	Cancer at t=1-2 years	-0.233	0.348	-0.127	0.606
	Cancer at t= 3-5 years	-0.601	0.030	-0.492	0.070
	Cancer at t=6-10 years	-0.042	0.821	0.033	0.850
	Cancer at t>10 years	-0.003	0.983	0.069	0.619
SF12 physical score	No cancer diagnosis	1	Reference	0	Reference
	Cancer at t=1-2 years	-10.229	<0.001	-8.454	<0.001
	Cancer at t=3-5 years	-9.288	<0.001	-7.762	<0.001
	Cancer at t=6-10 years	-6.441	<0.001	-5.349	<0.001
	Cancer at t>10 years	-5.147	<0.001	-4.062	<0.001
SF12 mental score	No cancer diagnosis	1	Reference	0	Reference
	Cancer at t=1-2 years	-1.758	0.088	-1.913	0.050
	Cancer at t=3-5 years	-0.622	0.501	-0.463	0.613
	Cancer at t=6-10 years	-0.156	0.844	-0.287	0.701
	Cancer at t>10 years	-3.139	<0.001	-2.981	<0.001

Notes: SF-12 = Short-Form 12, OLS = ordinary least squares. CI = confidence interval. Weights taken from UKHLS sample data.

E Additional Results for Analysis B in Chapter 6

Figure E.1.: Unmatched trajectories of proportions working in the cancer and non-cancer cohorts in Analysis B

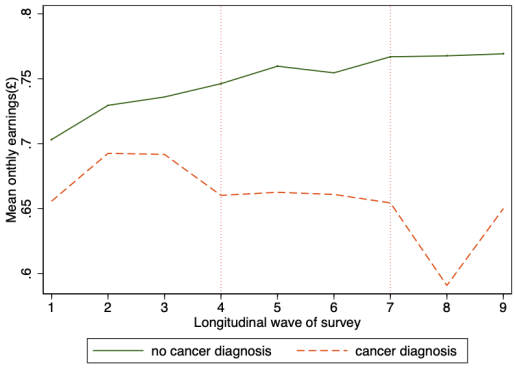


Figure E.2.: Unmatched trajectories of weekly hours worked in the cancer and non-cancer cohorts in Analysis B

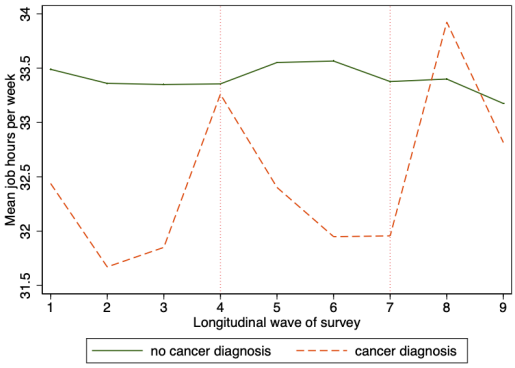


Figure E.3.: Unmatched trajectories of monthly earnings in the cancer and non-cancer cohorts in Analysis B

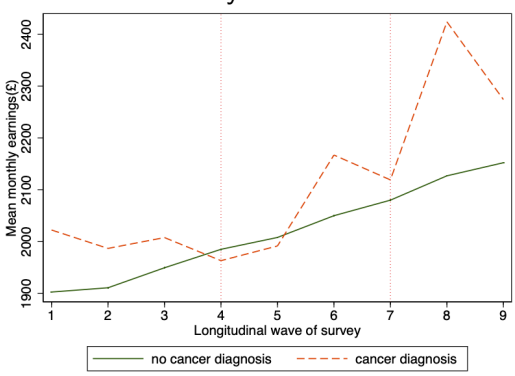


Figure E.4.: Unmatched trajectories of monthly income in the cancer and non-cancer cohorts in Analysis B

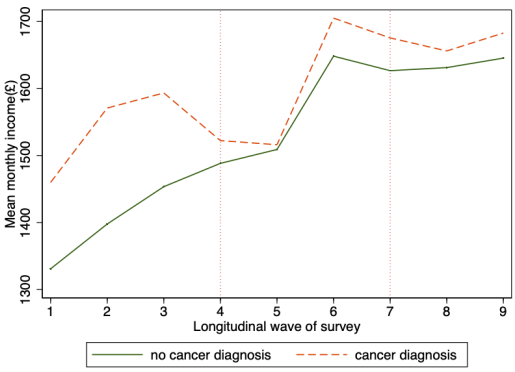


Figure E.5.: Unmatched trajectories of SF-12 physical score in the cancer and non-cancer cohorts in Analysis B

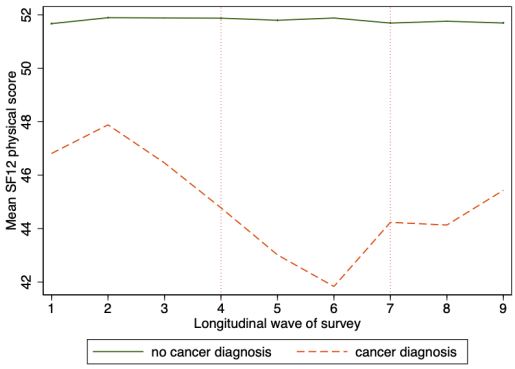


Figure E.6.: Unmatched trajectories of SF-12 mental score in the cancer and non-cancer cohorts in Analysis B